Implicit-Explicit Runge-Kutta schemes and finite elements with symmetric stabilization for advection-diffusion equations

Erik Burman, Alexandre Ern

Mathematical Modelling and Numerical Analysis Will be set by the publisher
Modélisation Mathématique et Analyse Numérique

IMPLICIT-EXPLICIT RUNGE–KUTTA SCHEMES AND FINITE ELEMENTS WITH SYMMETRIC STABILIZATION FOR ADVECTION–DIFFUSION EQUATIONS

Erik Burman[1] and Alexandre Ern[2]

Abstract. We analyze a two-stage explicit-implicit Runge–Kutta scheme for time discretization of advection–diffusion equations. Space discretization uses continuous, piecewise affine finite elements with interelement gradient jump penalty; discontinuous Galerkin methods can be considered as well. The advective and stabilization operators are treated explicitly, whereas the diffusion operator is treated implicitly. Our analysis hinges on L^2-energy estimates on discrete functions in physical space. Our main results are stability and quasi-optimal error estimates for smooth solutions under a standard hyperbolic CFL restriction on the time step, both in the advection-dominated and in the diffusion-dominated regimes. The theory is illustrated by numerical examples.

1991 Mathematics Subject Classification. 65M12, 65M15, 65M60.

Draft version: October 28, 2010.

1. Introduction

v_diff

We consider the transient advection–diffusion equation

$$\partial_t u + Bu + Au = f \quad \text{in } \Omega \times (0, t_{\mathrm{F}}), \tag{1a}$$

$$u = 0 \quad \text{on } \partial\Omega \times (0, t_{\mathrm{F}}), \tag{1b}$$

$$u(\cdot, t = 0) = u_0 \quad \text{in } \Omega, \tag{1c}$$

where Ω is a polyhedron in $\mathbb{R}^d$ with boundary $\partial\Omega$, $Bu := \beta\cdot\nabla u$, $Au := -\mu\Delta u$, t_{F} a finite positive time, β a divergence-free velocity field, $\mu > 0$ the diffusion coefficient, f the source term, and u_0 the initial datum. Extensions of the present analysis to advection fields with nonzero divergence and inclusion of non-stiff zero-order terms is straightforward; accounting for smoothly variable diffusion coefficient is also feasible.

In the stationary case, it is well-known that the standard Galerkin finite element method has poor stability properties in the advection-dominated regime, resulting in suboptimal convergence for smooth solutions and spurious oscillations when approximating solutions with sharp layers. Different approaches have been proposed to improve this behavior, such as the streamline upwind Petrov–Galerkin method (SUPG) [4, 24] and standard Galerkin methods with symmetric stabilization in various flavors, e.g., discontinuous Galerkin (DG) [17, 19, 25, 26], subgrid viscosity [20, 21], orthogonal subscale stabilization [14, 15], local projection stabilization [3, 29], and

Keywords and phrases: stabilized finite elements, stability, error bounds, implicit-explicit schemes, time-dependent PDEs

[1] Department of Mathematics, University of Sussex, Brighton, BN1 9RF United Kingdom; e-mail: E.N.Burman@sussex.ac.uk

[2] Université Paris-Est, CERMICS, Ecole des Ponts, 77455 Marne la Vallée Cedex 2, France; e-mail: ern@cermics.enpc.fr

continuous interior penalty on interelement normal gradient jumps (CIP) [5, 11]. All these methods lead to similar L^2-norm error estimates for smooth solutions, resulting in the loss of half a power of h in the advection-dominated regime (compared to a full power in the unstabilized case). For solutions with sharp layers, it has been proven for SUPG [24], DG [22], and CIP [10] that quasi-optimal convergence is retained away from layers, hence prohibiting the global spreading of spurious oscillations.

In the transient case, DG-based time discretization has been the favored alternative for SUPG [24], whereas Runge–Kutta (RK) methods have been popular for time discretization combined with DG in space [13]. For symmetric stabilizations in general, standard A-stable finite difference methods in time have been shown to be stable and optimally convergent [9, 15, 18, 21]. Similar results for SUPG and the transient advection–diffusion equation are very recent [6, 12]. The implicit time stepping by A-stable methods leads to a nonsymmetric matrix to be inverted at each time step. Moreover, treating nonlinear transport operators with such methods or incorporating nonlinear slope limiters can be quite demanding computationally. Ideally, one would like to treat the advective and stabilization operators explicitly and the diffusive operator implicitly. A suitable class of methods is that of implicit-explicit (IMEX) RK methods. The application of IMEX methods to partial differential equations (PDEs) was introduced in [16], and IMEX RK methods were first proposed in [1, 2]. From a computational viewpoint, IMEX RK methods only require symmetric systems to be solved at each time step, and the stencil of the corresponding matrix is that of the diffusion operator. Moreover, nonlinear transport operators and nonlinear slope limiters can be treated explicitly.

Although a substantial amount of literature exists on IMEX RK methods, deriving stability and error estimates for stabilized finite elements combined with IMEX RK time discretization remains, to the authors' knowledge, an open issue. In particular, we aim at an analysis that is valid in all flow regimes, that is, either advection-dominated or diffusion-dominated. Following the seminal work of Levy and Tadmor [27], the present analysis relies on L^2-energy estimates, that is, we work directly with discrete functions in the physical space. In other words, we account for the full geometric structure of eigenvectors, instead of the more classical approach using only scalar eigenvalue arguments which may be misleading in the context of nonnormal operators.

Concerning IMEX RK schemes, a first important issue is that the analysis of the truncation error in time by means of Butcher tables is not sufficient in the context of PDEs. In particular, this error involves the partial differential operators A and B acting on suitable functions associated with the intermediate stages of the scheme. In the IMEX scheme, bounding (high-order) derivatives of these functions is not straightforward and, in particular, requires a careful study of the role played by boundary conditions. A second important issue is that the explicit part of the RK scheme is anti-dissipative, that is, it produces energy, so that this energy production must be controlled by the stability induced by space discretization. In the context of finite element methods with symmetric stabilization, explicit (second- and third-order) RK methods were analyzed in [8], in particular for the pure advection equation, leading to stability and error estimates for smooth solutions. The presence of the diffusion operator poses additional difficulties to be tackled herein.

The two-stage IMEX RK scheme we consider for time discretization is the so-called SSP2(2,2,2) L-stable scheme proposed in [28] for hyperbolic systems with stiff relaxation terms and no sources. This scheme combines an explicit two-stage RK scheme for the transport operator together with a diagonally implicit, two-stage RK scheme for the stiff relaxation terms. Moreover, this scheme is formulated in terms of a parameter γ, and the value $\gamma = \gamma_* := 1 - \frac{1}{\sqrt{2}} \simeq 0.293$ is considered in [28]. Herein, we apply and analyze, for the first time, this scheme in the context of advection–diffusion equations. Space discretization is performed using continuous, piecewise affine finite elements with CIP as a specific example of symmetric stabilization; DG methods can be used as well, as discussed at the end of the manuscript. We treat the advection and stabilization operators explicitly and the diffusion operator implicitly.

Our main results are stability and error estimates for smooth solutions in all flow regimes. These results are formulated in terms of the Courant and Péclet numbers defined as

$$\mathrm{Co} := \frac{\sigma\tau}{h}, \qquad \mathrm{Pe} := \frac{\sigma h}{\mu},$$

where $\sigma := \|\beta\|_{L^\infty(\Omega)}$ is the reference velocity, h the mesh size, and τ the time step. For simplicity, the time step is taken to be constant, and we use a single Péclet number for the whole domain. In all flow regimes, we assume a hyperbolic type CFL restriction on the time step of the form $\mathrm{Co} \le \varrho$ with ϱ independent of the mesh size h, the time step τ, and the problem data. Furthermore, the analysis of the truncation error in time requires the technical assumptions that the normal component of β and the source term f vanish on $\partial\Omega$ and that elliptic regularity holds for the Laplace operator. In the advection-dominated regime ($\mathrm{Pe} \ge 1$), stability and convergence are achieved for $\gamma \in (0, \frac{1}{2})$; to fix the ideas, we take $\gamma \in [\frac{1}{5}, \frac{2}{5}]$ (the actual value of γ influences only the numerical bound on the Courant number). Our main convergence result (Theorem 4.2 and Proposition 4.1) takes the form

$$\|u(t_{\mathrm{F}}) - u_h^N\|_L + \left(\tau \sum_{n=1}^{N} \mu \|\nabla(u(t^n) - u_h^n)\|_{L^d}^2\right)^{1/2} \lesssim \tau^{3/2} + \sigma^{1/2} h^{3/2}.$$

The estimate for the space error is quasi-optimal (1/2-suboptimal), similarly to the steady case. The estimate for the time error is also quasi-optimal (1/2-suboptimal considering that a two-stage IMEX RK scheme is used). Owing to the CFL restriction on the time step, this estimate is actually sufficient to equilibrate space and time errors. In the diffusion-dominated regime ($\mathrm{Pe} \le 1$), stability and convergence are achieved for γ in a sufficiently small neighborhood of γ_*. In addition to the bound on the Courant number (which becomes trivial in the pure-diffusion limit), the time step is restricted by the bound $\tau \le (t_*/\mu)^{1/2} h$ where t_* is a reference time defined in §2.1. Our main convergence result (Theorem 4.3) takes the form

$$\left(\tau \sum_{n=1}^{N} \mu \|\nabla(u(t^n) - u_h^n)\|_{L^d}^2\right)^{1/2} \lesssim \tau + \mu^{1/2} h.$$

The estimate on the space error is optimal, while the estimate on the time error is 1-suboptimal, but, again, owing to the CFL restriction, it is actually sufficient to equilibrate space and time errors. Finally, still in the diffusion-dominated regime, we prove that (Proposition 4.2)

$$\|u(t_{\mathrm{F}}) - u_h^N\|_L \lesssim \tau^{3/2} + \sigma^{1/2} h^{3/2} + \mu^{-1/2} h^2.$$

This estimate is 1/2-suboptimal in time and in space, but, as the other estimates, equilibrates both errors owing to the CFL restriction. Moreover, as $\sigma \to 0$, that is, in the pure diffusion limit, second-order convergence is recovered in h. Finally, we observe that under an additional assumption on the boundary, the convergence order in time of all the above estimates can be improved by a factor $\tau^{1/2}$; see Remark 3.1.

The material is organized as follows. §2 states the basic assumptions, presents the setting for the space and time discretization, and introduces the truncation error in time together with the error equations. §3 is devoted to the analysis of the truncation error and the approximation error in space. §4 contains the stability and error analysis, while §5 presents numerical results. §6 discusses extensions to other space discretization schemes. In what follows, we often abbreviate $a \lesssim b$ the inequality $a \le Cb$ for positive C independent of the mesh size h, the time step τ, and the problem data. We only keep track of constants if they are to be used later in thresholds for the Courant number.

2. The setting

semi

In this section, we specify the basic assumptions for the time evolution problem (1) and the discretization parameters. We also present the stabilized finite element method for space discretization together with the two-stage IMEX RK scheme for time discretization. Then, we identify the truncation error in time upon introducing suitable intermediate functions associated with the intermediate stages of the IMEX RK scheme, and we derive the error equation. Finally, we collect important stability and boundedness properties of the discrete operators used for space discretization.

:basic

2.1. **Basic assumptions**

Let $L := L^2(\Omega)$ and let $V := H^2(\Omega) \cap H_0^1(\Omega)$. We assume that the exact solution u and the source term f are such that

$$u \in C^0([0,t_\mathrm{F}];H^4(\Omega)\cap H_0^1(\Omega)) \cap C^1([0,t_\mathrm{F}];H^3(\Omega)) \cap C^3([0,t_\mathrm{F}];L), \tag{2a}$$ eq:hyp.u

$$f \in C^0([0,t_\mathrm{F}];H^2(\Omega)\cap H_0^1(\Omega)) \cap C^2([0,t_\mathrm{F}];L), \tag{2b}$$ eq:hyp.f

and we observe that (2b) means, in particular, that $f|_{\partial\Omega} = 0$. We assume that the domain Ω is convex so that elliptic regularity holds true for the Laplace operator with homogeneous Dirichlet boundary conditions. Finally, we assume that β is in the Sobolev space $[W^{1,\infty}(\Omega)]^d$, so that β is bounded and has bounded derivatives, and that the normal component of β vanishes at the boundary, that is, $\nu\cdot\beta|_{\partial\Omega} = 0$ where ν denotes the unit outward normal to Ω. For later use, we set $\sigma_1 := \|\nabla\beta\|_{[L^\infty(\Omega)]^{d,d}}$ and observe that σ_1^{-1} can be interpreted as a time scale. We also consider the reference time $t_* := \min(\sigma_1^{-1}, t_\mathrm{F})$.

An important consequence of the fact that the normal component of β and the source term f vanish at the boundary is the following.

ABu.d0 **Proposition 2.1** (Boundary value of $Bu(t)$ and $Au(t)$)**.** *For all $t \in [0,t_\mathrm{F}]$,*

$$Bu(t)|_{\partial\Omega} = Au(t)|_{\partial\Omega} = 0. \tag{3}$$

Proof. The fact that $Bu(t)|_{\partial\Omega} = 0$ results from β having zero normal component on $\partial\Omega$ and u vanishing on $\partial\Omega$. The fact that $Au(t)|_{\partial\Omega} = 0$ then results from the evolution equation since $f(t)|_{\partial\Omega} = \partial_t u(t)|_{\partial\Omega} = 0$. □

Concerning the discretization parameters, we always assume to fix the ideas that $\mathrm{Co} \le 1$; bounds on the Courant number with different constants will be introduced later. We also assume the following mild reverse-parabolic CFL inequality

$$h^2 \lesssim \bar\mu\tau, \tag{4}$$ eq:reverse.C

where $\bar\mu := \max(\mu, \sigma^2 t_*)$. Finally, we make the mild assumption that the mesh size and the time step resolve the spatial variations of the advection velocity, that is,

$$\sigma_1 h \le \sigma, \qquad \sigma_1\tau \le 1, \tag{5}$$

and observe that the second bound implies $\tau \le t_*$ since $\tau \le t_\mathrm{F}$ as well.

2.2. **Space discretization**

Let $\{\mathcal{T}_h\}_{h>0}$ be a family of affine, simplicial meshes of Ω. We assume that the meshes are kept fixed in time and that the family $\{\mathcal{T}_h\}_{h>0}$ is quasi-uniform. It is also possible to work with shape-regular mesh families. In this case, as usual, the space scale in the CFL condition is no longer h, but the smallest element diameter in the mesh. Mesh faces are collected in the set $\mathcal{F}_h$ which is split into the set of interior faces, $\mathcal{F}_h^\mathrm{int}$, and boundary faces, $\mathcal{F}_h^\mathrm{ext}$. For a smooth enough function v that is possibly double-valued at $F \in \mathcal{F}_h^\mathrm{int}$ with $F = \partial T^- \cap \partial T^+$, we define its jump at F as $[\![v]\!] := v|_{T^-} - v|_{T^+}$, and we fix the unit normal vector to F, denoted by ν_F, as pointing from T^- to T^+. The arbitrariness in the sign of $[\![v]\!]$ is irrelevant in what follows.

Let V_h be the finite element space spanned by continuous and piecewise affine functions. Set $V(h) := H^2(\Omega) + V_h$. The space semi-discretized formulation can be written as follows: For all $t \in [0,t_\mathrm{F}]$, find $u_h(t) \in V_h$ such that

$$\partial_t u_h(t) + B_h u_h(t) + A_h u_h(t) = f_h(t), \tag{6}$$ space_semi_d

with initial condition $u_h(0) = \pi_h u_0$ and source term $f_h := \pi_h f$, where π_h denotes the L-orthogonal projection onto V_h. The discrete linear operators $B_h : V(h) \to V_h$ and $A_h : V(h) \to V_h$ are such that for all $(z, w_h) \in$

$V(h) \times V_h$,

$$(B_h z, w_h)_L := (\beta\cdot\nabla z, w_h)_L + \sum_{F\in\mathcal{F}_h^{\rm int}} S_{\rm cip} h_F^2 (|\nu_F\cdot\beta|\nu_F\cdot[\![\nabla z]\!], \nu_F\cdot[\![\nabla w_h]\!])_{L,F}, \tag{7a}$$

$$(A_h z, w_h)_L := (\mu\nabla z, \nabla w_h)_{L^d} - (\mu(\nu\cdot\nabla z), w_h)_{L,\partial\Omega} - (\mu z, \nu\cdot\nabla w_h)_{L,\partial\Omega} + S_{\rm bc} h^{-1} (\mu z, w_h)_{L,\partial\Omega}. \tag{7b}$$

Here, $(\cdot,\cdot)_L$ denotes the $L^2(\Omega)$-inner product (with associated norm $\|\cdot\|_L$) and $(\cdot,\cdot)_{L^d}$ the $[L^2(\Omega)]^d$-inner product (with associated norm $\|\cdot\|_{L^d}$), while for a subset $\omega\subset\Omega$ (a mesh face or a collection thereof), $(\cdot,\cdot)_{L,\omega}$ denotes the corresponding $L^2(\omega)$-inner product. We observe that the homogeneous Dirichlet boundary condition is weakly enforced in A_h (and that the additional boundary term $\sum_{F\in\mathcal{F}_h^{\rm ext}\cap\partial\Omega^-}(|\nu\cdot\beta|z, v_h)_{L,F}$, where $\partial\Omega^-$ denotes the inflow boundary, has been discarded from B_h since we assume $\nu\cdot\beta|_{\partial\Omega}=0$). Moreover, the user-dependent parameter $S_{\rm cip}$ is positive, while the user-dependent parameter $S_{\rm bc}$ is sufficiently large (see §2.6).

The discrete linear operators A_h and B_h satisfy important stability and boundedness properties collected in §2.6. For the time being, we record the following consistency property: For all $v\in V$,

$$B_h v = \pi_h(Bv), \qquad A_h v = \pi_h(Av). \tag{8}$$

2.3. **Time discretization**

For $0\le n\le N$ with $N:=\lfloor t_{\rm F}/\tau\rfloor$, a superscript n indicates the value of a function at the discrete time $n\tau$, and for $0\le n\le N-1$, we set $I_n:=(n\tau,(n+1)\tau]$. For a real parameter $\gamma\in(0,\frac12)$, we consider the following time discretization scheme:

eq:RK2

$$v_h^n = u_h^n - \gamma\tau A_h v_h^n + \gamma\tau f_h^n, \tag{9a}$$ eq:RK2.a

$$w_h^n = u_h^n - \tau B_h v_h^n - (1-2\gamma)\tau A_h v_h^n - \gamma\tau A_h w_h^n + (1-\gamma)\tau f_h^n, \tag{9b}$$ eq:RK2.b

$$u_h^{n+1} = u_h^n - \tfrac12\tau B_h(v_h^n + w_h^n) - \tfrac12\tau A_h(v_h^n + w_h^n) + \tau f_h^{n+\frac12}. \tag{9c}$$ eq:RK2.c

Here, $f_h^{n+\frac12}:=\pi_h f((n+\frac12)\tau)$ can be replaced by any second-order approximation in time, e.g., $\frac12(f_h^n+f_h^{n+1})$. We observe that the operator B_h is treated using an explicit two-stage RK scheme and the operator A_h using a diagonally implicit two-stage RK scheme. By using equation (9a) in (9b) and equations (9a)–(9b) in (9c), we obtain the following alternative form of the system (9):

K2.bis

$$v_h^n = u_h^n - \gamma\tau A_h v_h^n + \gamma\tau f_h^n, \tag{10a}$$ eq:RK2.aa

$$w_h^n = v_h^n - \tau B_h v_h^n - (1-3\gamma)\tau A_h v_h^n - \gamma\tau A_h w_h^n + (1-2\gamma)\tau f_h^n, \tag{10b}$$ eq:RK2.bb

$$u_h^{n+1} = \tfrac12(v_h^n + w_h^n) - \tfrac12\tau B_h w_h^n - \tfrac12\gamma\tau A_h v_h^n - \tfrac12(1-\gamma)\tau A_h w_h^n + \tau(f_h^{n+\frac12} - \tfrac12 f_h^n). \tag{10c}$$ eq:RK2.cc

2.4. **Truncation error in time**

The goal of this section is to identify the truncation error in time. Recalling the operators $B: V\ni v\mapsto \beta\cdot\nabla v\in L$ and $A: V\ni v\mapsto -\mu\Delta v\in L$, we introduce, for all $0\le n\le N-1$, the auxiliary functions $v^n, w^n\in H_0^1(\Omega)$ such that (compare with (9a)–(9b))

eq:rec

$$v^n + \gamma\tau A v^n = u^n + \gamma\tau f^n, \tag{11a}$$ eq:rec.v

$$w^n + \gamma\tau A w^n = u^n - \tau B v^n - (1-2\gamma)\tau A v^n + (1-\gamma)\tau f^n, \tag{11b}$$ eq:rec.w

or, equivalently, subtracting (11a) from (11b) (compare with (10b))

$$w^n + \gamma\tau A w^n = v^n - \tau B v^n - (1-3\gamma)\tau A v^n + (1-2\gamma)\tau f^n. \tag{12}$$ eq:rec.w.a

Moreover, owing to elliptic regularity, $v^n, w^n \in V$.

ef:Psi **Definition 2.1** (Truncation error)**.** *The truncation error $\Psi^n \in L$ at the discrete time $n\tau$ is defined as*

$$\Psi^n := \tau^{-1}(u^{n+1} - u^n) + \tfrac{1}{2}(A+B)(v^n + w^n) - f^{n+1/2}. \tag{13}$$ eq:def.Psi

It is straightforward to verify that (compare with (9c) and (10c))

$$\begin{aligned} u^{n+1} &= u^n - \tfrac{1}{2}\tau B(v^n + w^n) - \tfrac{1}{2}\tau A(v^n + w^n) + \tau f^{n+1/2} + \tau\Psi^n \\ &= \tfrac{1}{2}(v^n + w^n) - \tfrac{1}{2}\tau B w^n - \tfrac{1}{2}\gamma\tau A v^n - \tfrac{1}{2}(1-\gamma)\tau A w^n + \tau(f^{n+1/2} - \tfrac{1}{2}f^n) + \tau\Psi^n. \end{aligned} \tag{14}$$ eq:u.n+1

2.5. Error equation

To formulate the error equation, we define

$$\xi_h^n = u_h^n - \pi_h u^n, \qquad \theta_h^n = v_h^n - \pi_h v^n, \qquad \zeta_h^n = w_h^n - \pi_h w^n, \tag{15a}$$

$$\xi_\pi^n = u^n - \pi_h u^n, \qquad \theta_\pi^n = v^n - \pi_h v^n, \qquad \zeta_\pi^n = w^n - \pi_h w^n. \tag{15b}$$ eq:def.pi.n

Hence, the approximation error can be written as $u^n - u_h^n = -\xi_h^n + \xi_\pi^n$ and similarly for $v^n - v_h^n$ and $w^n - w_h^n$. The functions ξ_π^n, θ_π^n, and ζ_π^n are classically used to measure the space approximation errors.

q:pert **Lemma 2.1** (Error equation)**.** *There holds*

$$\theta_h^n = \xi_h^n - \gamma\tau A_h\theta_h^n + \tau\alpha_h^n, \tag{16a}$$ eq:pert.a

$$\zeta_h^n = \theta_h^n - \tau B_h\theta_h^n - (1-3\gamma)\tau A_h\theta_h^n - \gamma\tau A_h\zeta_h^n + \tau\beta_h^n, \tag{16b}$$ eq:pert.b

$$\xi_h^{n+1} = \tfrac{1}{2}(\theta_h^n + \zeta_h^n) - \tfrac{1}{2}\tau B_h\zeta_h^n - \tfrac{1}{2}\gamma\tau A_h\theta_h^n - \tfrac{1}{2}(1-\gamma)\tau A_h\zeta_h^n + \tau\delta_h^n - \tau\Psi_h^n, \tag{16c}$$ eq:pert.c

where $\Psi_h^n := \pi_h\Psi^n$ and

$$\alpha_h^n = \gamma A_h\theta_\pi^n, \qquad \beta_h^n = B_h\theta_\pi^n + (1-3\gamma)A_h\theta_\pi^n + \gamma A_h\zeta_\pi^n, \qquad \delta_h^n = \tfrac{1}{2}B_h\zeta_\pi^n + \tfrac{1}{2}\gamma A_h\theta_\pi^n + \tfrac{1}{2}(1-\gamma)A_h\zeta_\pi^n.$$

Proof. Apply the projector π_h to (11a), (12), and (14), use consistency, and subtract the resulting equations from (10). □

bnd.AB

2.6. Stability and boundedness of the discrete operators A_h and B_h

We define the following seminorm and norm on $V(h)$,

$$|z|_S^2 := \sum_{F\in\mathcal{F}_h^{\mathrm{int}}} S_{\mathrm{cip}} h_F^2 \| |\nu_F\cdot\beta|^{1/2}\nu_F\cdot[\![\nabla z]\!]\|_{L,F}^2, \tag{17a}$$ eq:def.S

$$\|z\|_A^2 := \mu\|\nabla z\|_{L^d}^2 + \mu h^{-1}\|z\|_{L,\partial\Omega}^2. \tag{17b}$$ eq:def.norme

It is well-known that provided S_{bc} is sufficiently large, there is $c_a > 0$ such that for all $v_h \in V_h$,

$$(A_h v_h, v_h)_L \geq c_a\|v_h\|_A^2. \tag{18}$$ eq:Ah.dissip

To allow for a more compact notation, we also consider the norm $\|v_h\|_a := (A_h v_h, v_h)_L^{1/2}$ for all $v_h \in V_h$. Furthermore, integration by parts readily yields

$$(B_h v_h, v_h)_L = |v_h|_S^2. \tag{19}$$ eq:Bh.dissip

We now examine briefly some important boundedness properties of the discrete operators A_h and B_h. In addition to the $|\cdot|_S$-seminorm and the $\|\cdot\|_A$-norm defined above, we consider the following norms on $V(h)$,

$$\|z\|_{B*} := |z|_S + \sigma^{1/2}h^{-1/2}\|z\|_L, \tag{20a}$$

$$\|z\|_{A*} := \|z\|_A + \mu^{1/2}h^{1/2}\|\nu\cdot\nabla z\|_{L,\partial\Omega}. \tag{20b}$$

These norms will be used to measure the space approximation errors. The following properties of B_h are established in [5, 7, 8].

Lemma 2.2 (Boundedness of B_h)**.** *For all $z \in V(h)$,*

$$\|B_h z\|_L \le \sigma\|\nabla z\|_{L^d} + C_S\sigma^{1/2}h^{-1/2}|z|_S, \tag{21}$$

for all $(z, v_h) \in V(h) \times V_h$,

$$|(B_h(z - \pi_h z), v_h)_L| \lesssim \|z - \pi_h z\|_{B*}(|v_h|_S + \sigma_1^{1/2}\|v_h\|_L), \tag{22}$$

and for all $(v_h, w_h) \in V_h \times V_h$,

$$|(B_h v_h, w_h - \pi_h^0 w_h)_L| \le C_B\sigma^{1/2}h^{-1/2}(|v_h|_S + \sigma_1^{1/2}\|v_h\|_L)\|w_h - \pi_h^0 w_h\|_L, \tag{23}$$

where π_h^0 denotes the L-orthogonal projection onto piecewise constant functions.

Using discrete trace and inverse inequalities, together with (21) yields for all $v_h \in V_h$,

$$|v_h|_S \lesssim \sigma^{1/2}h^{-1/2}\|v_h\|_L, \qquad \|B_h v_h\|_L \lesssim \sigma h^{-1}\|v_h\|_L, \tag{24}$$

while using (22) and the previous bound on $|v_h|_S$ yields for all $z \in V(h)$,

$$\tau\|B_h(z - \pi_h z)\|_L \lesssim \tau^{1/2}\mathrm{Co}^{1/2}\|z - \pi_h z\|_{B*}. \tag{25}$$

The following properties of A_h are established using fairly standard arguments, in particular discrete trace and inverse inequalities and the uniform equivalence of the $\|\cdot\|_A$- and $\|\cdot\|_{A*}$-norms on V_h.

Lemma 2.3 (Boundedness of A_h)**.** *For all $(z, w_h) \in V(h) \times V_h$,*

$$|(A_h z, w_h)_L| \lesssim \|z\|_{A*}\|w_h\|_A \quad \text{so that} \quad \|A_h z\|_L \lesssim \mu^{1/2}h^{-1}\|z\|_{A*}. \tag{26}$$

Additionally, for all $(z_h, w_h) \in V_h \times V_h$,

$$|(A_h z_h, w_h)_L| \lesssim \|z_h\|_A\|w_h\|_A \quad \text{so that} \quad \|A_h z_h\|_L \lesssim \mu^{1/2}h^{-1}\|z_h\|_A. \tag{27}$$

3. Truncation and space approximation errors

The goal of this section is to establish bounds on the truncation error Ψ^n defined by (13) and on the space approximation errors associated with the functions θ_π^n and ζ_π^n defined by (15b). To this end, we first derive bounds on the auxiliary functions at intermediate stages, namely the functions v^n and w^n defined by (11). Recall that owing to elliptic regularity, these functions are in $V = H^2(\Omega) \cap H_0^1(\Omega)$.

3.1. **Bounds on the auxiliary functions at intermediate stages**

Bounding Sobolev norms of the functions v^n and w^n hinges on the stability properties of the operator $(I+\gamma\tau A)$ (where I is the identity in V).

:basic **Lemma 3.1** (Stability of $(I+\gamma\tau A)$)**.** *Let $\mathfrak{v}\in L$ and let $\mathfrak{u}\in V$ be such that*

$$(I+\gamma\tau A)\mathfrak{u}=\mathfrak{v}. \tag{28}$$ eq:basic.eq

Then,

$$\|\mathfrak{u}\|_L\lesssim\|\mathfrak{v}\|_L,\qquad (\mu\tau)^{1/2}\|\nabla\mathfrak{u}\|_{L^d}\lesssim\|\mathfrak{v}\|_L. \tag{29}$$ eq:basic1

If, additionally $\mathfrak{v}\in H^1_0(\Omega)$,

$$\|\nabla\mathfrak{u}\|_{L^d}\lesssim\|\nabla\mathfrak{v}\|_{L^d},\qquad (\mu\tau)^{1/2}\|\Delta\mathfrak{u}\|_L\lesssim\|\nabla\mathfrak{v}\|_{L^d}. \tag{30}$$ eq:basic2

If, additionally $\mathfrak{v}\in V$,

$$\|\Delta\mathfrak{u}\|_L\lesssim\|\Delta\mathfrak{v}\|_L,\qquad (\mu\tau)^{1/2}\|\nabla\Delta\mathfrak{u}\|_{L^d}\lesssim\|\Delta\mathfrak{v}\|_L. \tag{31}$$ eq:basic3

Proof. Take the L-scalar product of (28) with $\mathfrak{u}$ and integrate by parts to infer (29), apply the same procedure to (28) with $\Delta\mathfrak{u}$ observing that $\Delta\mathfrak{u}|_{\partial\Omega}=0$ owing to (28) to infer (30), and take the Laplacian of (28) and apply the same procedure with $\Delta\mathfrak{u}$ to infer (31). □

As a first application, we derive bounds on (v^n-u^n) and on v^n.

nd.v-u **Lemma 3.2** (Bounds on (v^n-u^n) and v^n)**.** *For $s\in\{1,2\}$, set $K^n_s:=|f^n|_{H^s}+\mu|u^n|_{H^{s+2}}$. Then,*

$$\|\nabla(v^n-u^n)\|_{L^d}\lesssim\tau K^n_1,\qquad \|\Delta(v^n-u^n)\|_L\lesssim\tau K^n_2,\qquad (\mu\tau)^{1/2}\|\nabla\Delta(v^n-u^n)\|_L\lesssim\tau K^n_2, \tag{32}$$ eq:bnd.v-u

and letting $\tilde K^n_s=|u^n|_{H^s}+\tau K^n_s$,

$$\|\nabla v^n\|_{L^d}\lesssim\tilde K^n_1,\qquad |v^n|_{H^2}\lesssim\tilde K^n_2. \tag{33}$$ eq:bnd.v

Proof. Take $\mathfrak{u}:=v^n-u^n$ so that $\mathfrak{v}=\gamma\tau(f^n-Au^n)$ owing to (11a). Since $\mathfrak{v}\in V$ (recall that f^n and Au^n vanish on $\partial\Omega$), the bound on $\|\nabla(v^n-u^n)\|_{L^d}$ results from (30) and the two other bounds on (v^n-u^n) from (31). Finally, the bounds (33) on v^n result from (32), the triangle inequality, and elliptic regularity. □

As a second application, we derive bounds on (w^n-u^n) and on w^n.

nd.w-u **Lemma 3.3** (Bounds on (w^n-u^n) and w^n)**.** *Let $K^n_{w-u}:=K^n_1+\sigma\tilde K^n_2+\sigma_1\tilde K^n_1$. Then,*

$$\|\nabla(w^n-u^n)\|_{L^d}\lesssim\tau K^n_{w-u},\qquad (\mu\tau)^{1/2}\|\Delta(w^n-u^n)\|_L\lesssim\tau K^n_{w-u}, \tag{34}$$ eq:bnd.w-u

and

$$(\mu\tau)^{1/2}|w^n|_{H^2}\lesssim(\mu\tau)^{1/2}|u^n|_{H^2}+\tau K^n_{w-u}. \tag{35}$$ eq:bnd.w

Proof. We first deduce from (11) that

$$(I+\gamma\tau A)(w^n-u^n)=\gamma\tau(f^n-Au^n)+\gamma^{-1}(1-2\gamma)(v^n-u^n)-\tau Bv^n. \tag{36}$$ eq:fkv.w-u

As a result, we can apply Lemma 3.1 with $\mathfrak{u}:=w^n-u^n$ and $\mathfrak{v}$ equal to the right-hand side of (36). We observe that $\mathfrak{v}\in H^1_0(\Omega)$ and that $\|\nabla\mathfrak{v}\|_{L^d}\lesssim\tau K^n_{w-u}$ since, in particular, $\|\nabla(Bv^n)\|_{L^d}\le\sigma|v^n|_{H^2}+\sigma_1\|\nabla v^n\|_{L^d}\lesssim\sigma\tilde K^n_2+\sigma_1\tilde K^n_1$ where we have used (33) to bound v^n. Hence, the bounds (34) on (w^n-u^n) result from (30). Finally, the bound (35) on w^n results from (34), the triangle inequality, and elliptic regularity. □

3.2. **Bound on the truncation error**

In this section, we derive two bounds on the truncation error. To this end, it is useful to consider the following equivalent expression for Ψ^n (the proof, which amounts to a direct verification, is skipped for brevity).

em:Psi **Lemma 3.4** (Equivalent expression for Ψ^n)**.** *Let $x^n \in V$ be defined such that*

$$x^n := \tfrac{1}{2}(v^n + w^n) - u^n - \tfrac{1}{2}\tau\partial_t u^n. \tag{37}$$ eq:def.x

Then, letting $\tilde{\Psi}^n := \tau^{-1}(u^{n+1} - u^n - \tau\partial_t u^n - \frac{1}{2}\tau^2\partial_{tt}u^n) + (f^n + \frac{1}{2}\tau\partial_t f^n - f^{n+1/2})$, there holds

$$\Psi^n = \tilde{\Psi}^n + Bx^n + Ax^n. \tag{38}$$ eq:Psi.ABx

We observe that it is necessary to bound spatial derivatives of x^n in order to control the terms Bx^n and Ax^n. Here, the bounds on $(v^n - u^n)$ derived in Lemma 3.2 are instrumental.

:bnd.x nd:ABx **Lemma 3.5** (Bounds on x^n)**.** *Let $C_x^n := \mu^{1/2}K_2^n + \tau^{1/2}(\sigma K_2^n + \sigma_1 K_1^n + \mu|\partial_t u^n|_{H^3})$. Then,*

$$\|Bx^n\|_L \lesssim \sigma C_x^n \tau^{3/2}, \tag{39a}$$ eq:bnd.Bx
$$\|Ax^n\|_L \lesssim \mu^{1/2} C_x^n \tau, \tag{39b}$$ eq:bnd.Ax
$$\|x^n\|_{A*} \lesssim \bar{\mu}^{1/2} C_x^n \tau^{3/2}. \tag{39c}$$ eq:bnd.xA

Proof. A direct calculation shows that

$$y^n := (I + \gamma\tau A)x^n = -\tfrac{1}{2}\tau B(v^n - u^n) - \tfrac{1}{2}(1 - 2\gamma)\tau A(v^n - u^n) - \tfrac{1}{2}\gamma\tau^2 A\partial_t u^n. \tag{40}$$ eq:x.y

Applying Lemma 3.1 with $\mathfrak{u} = x^n$ and $\mathfrak{v} = y^n$ and observing that $y^n \in H_0^1(\Omega)$ (for the first term, $\nu\cdot\beta$ as well as $(v^n - u^n)$ vanish on $\partial\Omega$; for the second term, Av^n vanishes on $\partial\Omega$ owing to (11a) and Au^n by Proposition 2.1; for the third term, $Au(t)$ vanishes on $\partial\Omega$ at all times by Proposition 2.1 and, hence, so does its time-derivative), we infer using (30) that $\|\nabla x^n\|_{L^d} \lesssim \|\nabla y^n\|_{L^d}$ and $(\mu\tau)^{1/2}\|\Delta x^n\|_L \lesssim \|\nabla y^n\|_{L^d}$. Using the bounds (32) on $(v^n - u^n)$ yields $\|\nabla y^n\|_{L^d} \lesssim C_x^n\tau^{3/2}$, whence (39a) and (39b). Finally, a continuous scaled trace inequality together with elliptic regularity yield

$$\|x^n\|_{A*} \lesssim \mu^{1/2}(\|\nabla x^n\|_{L^d} + h|x^n|_{H^2}) \lesssim \mu^{1/2}(\|\nabla x^n\|_{L^d} + h\|\Delta x^n\|_L).$$

Using the reverse-parabolic CFL inequality (4) and the above bounds on $\|\nabla x^n\|_{L^d}$ and $\|\Delta x^n\|_L$, we infer

$$\|x^n\|_{A*} \lesssim \bar{\mu}^{1/2}(\|\nabla x^n\|_{L^d} + (\mu\tau)^{1/2}\|\Delta x^n\|_L) \lesssim \bar{\mu}^{1/2}\|\nabla y^n\|_{L^d},$$

whence (39c) results from the bound on $\|\nabla y^n\|_{L^d}$. □

We can now state the main result of this section, providing two ways to bound the truncation error. The first bound (42a) is simpler, but is only first-order in time; the second bound (42b) is of higher-order, namely 3/2, but estimates the diffusive contribution of x^n differently. Both bounds will be used in what follows.

nd.Psi **Lemma 3.6.** *Let*

$$C_\Psi^n := (t_*\tau)^{1/2} C_{u,f}^n + t_*^{1/2}\sigma C_x^n + \bar{\mu}^{1/2} C_x^n, \tag{41a}$$ eq:Cpsi
$$\tilde{C}_\Psi^n := \tau C_{u,f}^n + \tau^{1/2}\sigma C_x^n + \bar{\mu}^{1/2} C_x^n, \tag{41b}$$ eq:tCpsi

where C_x^n *is defined in Lemma* 3.5 *and* $C_{u,f}^n := \|u\|_{C^3(I_n;L)} + \|f\|_{C^2(I_n;L)}$. *Then,*

$$\|\Psi^n\|_L \le \|\tilde\Psi^n\|_L + \|Bx^n\|_L + \|Ax^n\|_L \lesssim \tilde C_\Psi^n \tau, \tag{42a}$$ eq:bnd.Psi

$$\|\tilde\Psi^n\|_L + \|Bx^n\|_L + t_*^{-1/2}\|x^n\|_{A*} \lesssim t_*^{-1/2} C_\Psi^n \tau^{3/2}. \tag{42b}$$ eq:bnd.Psi.s

Proof. Using the definition (38) and the triangle inequality leads to

$$\|\Psi^n\|_L \le \|\tilde\Psi^n\|_L + \|Bx^n\|_L + \|Ax^n\|_L,$$

whence (42a) results from (39a), (39b), and the obvious bound $\|\tilde\Psi^n\|_L \le C_{u,f}^n \tau^2$. Furthermore, the second bound (42b) results from (39a), (39c), and the same bound on $\|\tilde\Psi^n\|_L$. □

r.time *Remark* 3.1 (Convergence order in time). Although the two-stage IMEX RK scheme is formally of second-order, as reflected by the bound on $\|\tilde\Psi^n\|_L$ based on Taylor polynomial expansions on u and f, the bounds on the truncation error derived in Lemma 3.6 are not of second-order. In fact, although $\|x^n\|_L$ is second-order in time (this results from (40) so that $\|x^n\|_L \le \|y^n\|_L$ and the fact that $\|y^n\|_L \lesssim \tau^2(\sigma K_1 + \mu K_2 + \mu|\partial_t u^n|_{H^2})$), the first- and second-order derivatives of x^n are not second-order in time, as reflected by the bounds derived in Lemma 3.5 on $\|Bx^n\|_L$ and $\|Ax^n\|_L$. The difficulty in deriving higher-order bounds on $\|Bx^n\|_L$ and $\|Ax^n\|_L$ stems from boundary conditions. To establish the present bounds, we have, in particular, made use of $Au^n|_{\partial\Omega} = 0$ and $Bu^n|_{\partial\Omega} = 0$ owing to Proposition 2.1. Under the more restrictive assumption $ABu^n|_{\partial\Omega} = 0$ (which holds true, e.g., if the normal derivative of β and the Laplacian of the normal component of β vanish on $\partial\Omega$), it is possible to gain a factor $\tau^{1/2}$ in the bounds on $\|Bx^n\|_L$ and $\|Ax^n\|_L$. This results from the fact that the function y^n defined by (40) is such that $(I+\gamma\tau A)y^n = \tau^2(z_1^n + z_2^n)$ with $z_1^n = -\frac12\gamma B(f^n - Au^n) - \frac12(1-2\gamma)\gamma A(f^n - Au^n) - \frac12\gamma A\partial_t u^n \in H_0^1(\Omega)$ (since $ABu^n|_{\partial\Omega} = 0$) and $z_2^n = -\frac12\gamma(AB - BA)(v^n - u^n) - \frac12\gamma^2\tau A^2\partial_t u^n \in L$ so that $\|\nabla y^n\|_{L^d} \lesssim \tau^2$ (details are skipped for brevity). An alternative assumption leading to the same conclusion is to use periodic boundary conditions. Finally, we stress that the present bounds are, however, sufficient to equilibrate the space and time errors in our error estimates in the context of the CFL restriction on the time step.

3.3. **Bounds on the space approximation errors**

The goal of this section is to bound the $\|\cdot\|_{A*}$- and $\|\cdot\|_{B*}$-norms of θ_π^n and ζ_π^n. We first observe that standard approximation properties in finite element spaces yield for all $z \in H^2(\Omega)$,

$$\|z - \pi_h z\|_{B*} \lesssim \sigma^{1/2} h^{3/2} |z|_{H^2}, \qquad \|z - \pi_h z\|_{A*} \lesssim \mu^{1/2} h |z|_{H^2}. \tag{43}$$ eq:app.star

th.zet **Lemma 3.7** (Bound on θ_π^n and ζ_π^n)**.** *There holds*

$$\|\theta_\pi^n\|_{B*} + \|\theta_\pi^n\|_{A*} \lesssim (\sigma^{1/2}h^{3/2} + \mu^{1/2}h)\tilde K_2^n, \tag{44a}$$ eq:bnd.theta

$$\|\zeta_\pi^n\|_{B*} + \|\zeta_\pi^n\|_{A*} \lesssim (\sigma^{1/2}h^{3/2} + \mu^{1/2}h)\tilde K_2^n + \tau^{1/2} h K_{w-u}^n. \tag{44b}$$ eq:bnd.zeta.

Proof. The bound (44a) readily results from (43) and the bound (33) on $|v^n|_{H^2}$. To bound $\|\zeta_\pi^n\|_{A*}$, we use again (43) together with (35) yielding $\|\zeta_\pi^n\|_{A*} \lesssim \mu^{1/2}h|u^n|_{H^2} + \tau^{1/2}hK_{w-u}^n$. To bound $\|\zeta_\pi^n\|_{B*}$, we first observe that for a function $z \in V$,

$$\|z - \pi_h z\|_{B*} \lesssim \sigma^{1/2} h^{1/2} \|\nabla z\|_{L^d}.$$

This assertion is clear for the $\|\cdot\|_L$-norm contribution, while using a discrete trace inequality and the H^1-stability of π_h yields

$$|z - \pi_h z|_S = |\pi_h z|_S \lesssim \sigma^{1/2} h^{1/2} \|\nabla \pi_h z\|_{L^d} \lesssim \sigma^{1/2} h^{1/2} \|\nabla z\|_{L^d}. \tag{45}$$ eq:bnd.z-Piz

As a result, starting from the triangle inequality

$$\|\zeta_\pi^n\|_{B*} \le \|u^n - \pi_h u^n\|_{B*} + \|(w^n - u^n) - \pi_h(w^n - u^n)\|_{B*},$$

and using the approximation property (43) for the first term together with (45), we infer

$$\|\zeta_\pi^n\|_{B*} \lesssim \sigma^{1/2}h^{3/2}|u^n|_{H^2} + \sigma^{1/2}h^{1/2}\tau K_{w-u}^n \leq \sigma^{1/2}h^{3/2}|u^n|_{H^2} + \tau^{1/2}hK_{w-u}^n,$$

where we have used (34) to bound $\|\nabla(w^n - u^n)\|_{L^d}$ and the fact that $\mathrm{Co} \leq 1$. The conclusion is straightforward since $|u^n|_{H^2} \leq \tilde{K}_2^n$. □

CV.RK2

4. Stability and convergence analysis

This section is devoted to the stability and convergence analysis of the IMEX RK scheme (10). Firstly, we derive a basic energy estimate valid in all flow regimes (Theorem 4.1). On the right-hand side of this estimate appears an anti-dissipative term together with the time and space discretization errors. Then, we bound the anti-dissipative term depending on the flow regime, yielding our main convergence results (Theorems 4.2 and 4.3 together with Propositions 4.1 and 4.2).

4.1. Basic energy identity

We begin the analysis with a basic energy identity valid in all flow regimes.

rgy.id

Lemma 4.1 (Basic energy identity)**.** *Assume $\gamma \in (0, \frac{1}{2})$. There holds*

$$\begin{aligned}&\tfrac{1}{2}\|\xi_h^{n+1}\|_L^2 - \tfrac{1}{2}\|\xi_h^n\|_L^2 + \tfrac{1}{2}\|\theta_h^n - \xi_h^n\|_L^2 + \tfrac{1}{2}\tau|\theta_h^n|_S^2 + \tfrac{1}{2}\tau|\zeta_h^n|_S^2 + \left(\tfrac{1}{2}-\gamma\right)\tau\|\theta_h^n\|_a^2 + \left(\tfrac{1}{2}-\gamma\right)\tau\|\zeta_h^n\|_a^2 \\ &+ \tfrac{1}{2}\gamma\tau\|\zeta_h^n + \theta_h^n\|_a^2 = \tfrac{1}{2}\|\xi_h^{n+1} - \zeta_h^n\|_L^2 + \tau(\alpha_h^n + \tfrac{1}{2}\beta_h^n, \theta_h^n)_L + \tau(\delta_h^n, \zeta_h^n)_L - \tau(\Psi_h^n, \zeta_h^n)_L.\end{aligned} \tag{46}$$

energy.id

Remark 4.1 (Pure advection, role of diffusion). Setting the diffusion coefficient to zero, the energy identity (46) reduces to the one derived in [8] for explicit RK2 schemes in the purely advective case. Moreover, in the presence of diffusion, all the additional terms involving the $\|\cdot\|_a$-norm are dissipative for $\gamma \in (0, \frac{1}{2})$.

Proof. We multiply equation (16a) by θ_h^n to obtain using the discrete stability (18) of A_h,

$$\tfrac{1}{2}\|\theta_h^n\|_L^2 + \tfrac{1}{2}\|\theta_h^n - \xi_h^n\|_L^2 = \tfrac{1}{2}\|\xi_h^n\|_L^2 + (\theta_h^n - \xi_h^n, \theta_h^n)_L = \tfrac{1}{2}\|\xi_h^n\|_L^2 - \gamma\tau\|\theta_h^n\|_a^2 + \tau(\alpha_h^n, \theta_h^n)_L. \tag{47}$$

stab1

Then, we multiply equation (16b) by $\frac{1}{2}\theta_h^n$ and equation (16c) by ζ_h^n to obtain

$$\tfrac{1}{2}(\zeta_h^n, \theta_h^n)_L = \tfrac{1}{2}\|\theta_h^n\|_L^2 - \tfrac{1}{2}\tau(B_h\theta_h^n, \theta_h^n)_L - \tfrac{1}{2}(1-3\gamma)\tau\|\theta_h^n\|_a^2 - \tfrac{1}{2}\gamma\tau(A_h\zeta_h^n, \theta_h^n)_L + \tfrac{1}{2}\tau(\beta_h^n, \theta_h^n)_L \tag{48}$$

stab2

and

$$(\xi_h^{n+1}, \zeta_h^n)_L = \tfrac{1}{2}(\theta_h^n + \zeta_h^n, \zeta_h^n)_L - \tfrac{1}{2}\tau(B_h\zeta_h^n, \zeta_h^n)_L - \tfrac{1}{2}\gamma\tau(A_h\theta_h^n, \zeta_h^n)_L - \tfrac{1}{2}(1-\gamma)\tau\|\zeta_h^n\|_a^2 + \tau(\delta_h^n - \Psi_h^n, \zeta_h^n)_L. \tag{49}$$

stab3

Summing (47) and (48) we deduce

$$\begin{aligned}\tfrac{1}{2}(\zeta_h^n, \theta_h^n)_L = &-\tfrac{1}{2}\|\theta_h^n - \xi_h^n\|_L^2 + \tfrac{1}{2}\|\xi_h^n\|_L^2 - \tfrac{1}{2}\tau(B_h\theta_h^n, \theta_h^n)_L - \tfrac{1}{2}(1-\gamma)\tau\|\theta_h^n\|_a^2 - \tfrac{1}{2}\gamma\tau(A_h\zeta_h^n, \theta_h^n)_L \\ &+ \tau(\alpha_h^n + \tfrac{1}{2}\beta_h^n, \theta_h^n)_L.\end{aligned} \tag{50}$$

stab4

Using now the identity $(\xi_h^{n+1}, \zeta_h^n)_L = \frac{1}{2}\|\xi_h^{n+1}\|_L^2 - \frac{1}{2}\|\xi_h^{n+1} - \zeta_h^n\|_L^2 + \frac{1}{2}\|\zeta_h^n\|_L^2$ together with (49) and (50), we infer

$$\begin{aligned}&\tfrac{1}{2}\|\xi_h^{n+1}\|_L^2 - \tfrac{1}{2}\|\xi_h^{n+1} - \zeta_h^n\|_L^2 + \tfrac{1}{2}\|\zeta_h^n\|_L^2 = \tfrac{1}{2}\|\zeta_h^n\|_L^2 - \tfrac{1}{2}\|\theta_h^n - \xi_h^n\|_L^2 + \tfrac{1}{2}\|\xi_h^n\|_L^2 - \tfrac{1}{2}\tau(B_h\theta_h^n, \theta_h^n)_L - \tfrac{1}{2}\tau(B_h\zeta_h^n, \zeta_h^n)_L \\ &- \tfrac{1}{2}(1-\gamma)\tau\|\theta_h^n\|_a^2 - \gamma\tau(A_h\zeta_h^n, \theta_h^n)_L - \tfrac{1}{2}(1-\gamma)\tau\|\zeta_h^n\|_a^2 + \tau(\alpha_h^n + \tfrac{1}{2}\beta_h^n, \theta_h^n)_L + \tau(\delta_h^n, \zeta_h^n)_L - \tau(\Psi_h^n, \zeta_h^n)_L.\end{aligned}$$

Rearranging the relation, completing the square in the three terms involving the $\|\cdot\|_a$-norm, and using the discrete stability (19) of B_h yields the assertion. □

4.2. **Bound on source terms and basic energy estimate**

The goal of our second step is to bound the contributions of the source terms α_h^n, β_h^n, δ_h^n, and Ψ_h^n on the right-hand side of the basic energy identity (46). To this purpose, we exploit the presence of the $|\cdot|_S^2$-terms and the $\|\cdot\|_a^2$-terms on the left-hand side (Lemma 4.2) so as to arrive at a basic energy estimate valid in all flow regimes and where the only term left to be bounded is the anti-dissipative term $\frac{1}{2}\|\xi_h^{n+1}-\zeta_h^n\|_L^2$ (Theorem 4.1). To fix the ideas, we assume $\gamma\in[\frac{1}{5},\frac{2}{5}]$. A larger interval included in $(0,\frac{1}{2})$ can be considered; this will only modify the numerical factors in front of the $\|\cdot\|_a^2$-terms. We introduce the quantity

$$E_h^n := t_*^{-1/2}\|\xi_h^n\|_L + \|\theta_\pi^n\|_{B*} + \|\theta_\pi^n\|_{A*} + \|\zeta_\pi^n\|_{B*} + \|\zeta_\pi^n\|_{A*} + C_\Psi^n\tau^{3/2}, \tag{51}$$ eq:def.Ehn

which collects, in addition to $t_*^{-1/2}\|\xi_h^n\|_L$, the space and time approximation errors. The contribution of the truncation error is already bounded in terms of the time step and the constant C_Ψ^n defined by (41a); instead, we do not yet bound the space approximation errors to keep track of these quantities in the proofs below.

source **Lemma 4.2** (Bound on the source terms)**.** *Assume* $\gamma\in[\frac{1}{5},\frac{2}{5}]$ *and* $\mathrm{Co}\le 1$. *Then,*

$$\tau|(\alpha_h^n+\tfrac{1}{2}\beta_h^n,\theta_h^n)_L + (\delta_h^n,\zeta_h^n)_L - (\Psi_h^n,\zeta_h^n)_L| \le \tfrac{1}{8}\tau|\theta_h^n|_S^2 + \tfrac{1}{8}\tau|\zeta_h^n|_S^2 + \tfrac{1}{40}\tau\|\theta_h^n\|_a^2 + \tfrac{1}{80}\tau\|\zeta_h^n\|_a^2 + C\tau(E_h^n)^2. \tag{52}$$ eq:bnd.sourc

Proof. We first bound $\|\theta_h^n\|_L$ and $\|\zeta_h^n\|_L$. Taking the L-scalar product of (16a) with θ_h^n yields

$$\|\theta_h^n\|_L^2 + \gamma\tau\|\theta_h^n\|_a^2 = (\xi_h^n,\theta_h^n)_L + \gamma\tau(A_h\theta_\pi^n,\theta_h^n)_L.$$

Using (26) and the Cauchy–Schwarz inequality yields $\|\theta_h^n\|_L^2+\tau\|\theta_h^n\|_a^2 \lesssim \|\xi_h^n\|_L\|\theta_h^n\|_L + \tau\|\theta_\pi^n\|_{A*}\|\theta_h^n\|_A$. Hence, using Young's inequality together with (18), we obtain

$$\|\theta_h^n\|_L^2 + \tau\|\theta_h^n\|_a^2 \lesssim \|\xi_h^n\|_L^2 + \tau\|\theta_\pi^n\|_{A*}^2. \tag{53}$$ eq:bnd.theta

Taking now the L-scalar product of (16b) with ζ_h^n yields

$$\|\zeta_h^n\|_L^2 + \gamma\tau\|\zeta_h^n\|_a^2 = (\theta_h^n,\zeta_h^n)_L - \tau(B_h\theta_h^n,\zeta_h^n)_L - (1-3\gamma)\tau(A_h\theta_h^n,\zeta_h^n)_L + \tau(\beta_h^n,\zeta_h^n)_L.$$

Using (24), the Cauchy–Schwarz inequality, and $\mathrm{Co}\le 1$, we infer $\tau|(B_h\theta_h^n,\zeta_h^n)_L| \lesssim \|\theta_h^n\|_L\|\zeta_h^n\|_L$. In addition, $\tau|(A_h\theta_h^n,\zeta_h^n)_L| \lesssim \tau\|\theta_h^n\|_a\|\zeta_h^n\|_a$ owing to (27) and (18), while using the boundedness (22) and (26) of B_h and A_h, we infer

$$\begin{aligned}\tau|(\beta_h^n,\zeta_h^n)_L| &\lesssim \tau\|\theta_\pi^n\|_{B*}(|\zeta_h^n|_S + \sigma_1^{1/2}\|\zeta_h^n\|_L) + \tau(\|\theta_\pi^n\|_{A*}+\|\zeta_\pi^n\|_{A*})\|\zeta_h^n\|_A\\ &\lesssim \tau^{1/2}\|\theta_\pi^n\|_{B*}\|\zeta_h^n\|_L + \tau(\|\theta_\pi^n\|_{A*}+\|\zeta_\pi^n\|_{A*})\|\zeta_h^n\|_A,\end{aligned}$$

where we have used $\tau\sigma_1\le 1$, $\mathrm{Co}\le 1$, and (24). Hence,

$$\|\zeta_h^n\|_L^2 + \tau\|\zeta_h^n\|_a^2 \lesssim \|\theta_h^n\|_L^2 + \tau\|\theta_h^n\|_a^2 + \tau(\|\theta_\pi^n\|_{B*}^2 + \|\theta_\pi^n\|_{A*}^2 + \|\zeta_\pi^n\|_{A*}^2),$$

and accounting for (53) finally yields

$$\|\zeta_h^n\|_L^2 + \tau\|\zeta_h^n\|_a^2 \lesssim \|\xi_h^n\|_L^2 + \tau(\|\theta_\pi^n\|_{B*}^2 + \|\theta_\pi^n\|_{A*}^2 + \|\zeta_\pi^n\|_{A*}^2). \tag{54}$$ eq:bnd.zeta

We are now ready to bound the source terms. Since $\alpha_h^n = \gamma A_h\theta_\pi^n$ and $|(A_h\theta_\pi^n,\theta_h^n)_L| \lesssim \|\theta_\pi^n\|_{A*}\|\theta_h^n\|_A \lesssim \|\theta_\pi^n\|_{A*}\|\theta_h^n\|_a$ owing to (26) and (18), we first obtain using Young's inequality

$$\tau|(\alpha_h^n,\theta_h^n)_L| \le \tfrac{1}{80}\tau\|\theta_h^n\|_a^2 + C\tau\|\theta_\pi^n\|_{A*}^2 \le \tfrac{1}{80}\tau\|\theta_h^n\|_a^2 + C\tau(E_h^n)^2. \tag{55}$$ eq:bnd.aa

Similarly, recalling $\beta_h^n = B_h\theta_\pi^n + (1-3\gamma)A_h\theta_\pi^n + \gamma A_h\zeta_\pi^n$ and using (22),

$$\tfrac{1}{2}\tau|(\beta_h^n,\theta_h^n)_L| \le \tfrac{1}{8}\tau(|\theta_h^n|_S^2 + \sigma_1\|\theta_h^n\|_L^2) + \tfrac{1}{80}\tau\|\theta_h^n\|_a^2 + C\tau(\|\theta_\pi^n\|_{B*}^2 + \|\theta_\pi^n\|_{A*}^2 + \|\zeta_\pi^n\|_{A*}^2).$$

Hence, using (53) to bound $\|\theta_h^n\|_L$ and since $\tau \le t_* \le \sigma_1^{-1}$, we infer

$$\tfrac{1}{2}\tau|(\beta_h^n,\theta_h^n)_L| \le \tfrac{1}{8}\tau|\theta_h^n|_S^2 + \tfrac{1}{80}\tau\|\theta_h^n\|_a^2 + C\tau(E_h^n)^2. \tag{56}$$ eq:bnd.bb

Turning to δ_h^n and recalling that $\delta_h^n = \frac{1}{2}B_h\zeta_\pi^n + \frac{1}{2}\gamma A_h\theta_\pi^n + \frac{1}{2}(1-\gamma)A_h\zeta_\pi^n$ and proceeding as above, we infer

$$\tau|(\delta_h^n,\zeta_h^n)_L| \le \tfrac{1}{8}\tau|\zeta_h^n|_S^2 + \tfrac{1}{160}\tau\|\zeta_h^n\|_a^2 + C\tau(E_h^n)^2. \tag{57}$$ eq:bnd.dd

Finally, concerning Ψ_h^n, we infer using (38), the Cauchy–Schwarz inequality, and Young's inequality (note in particular that $(Ax^n,\zeta_h^n)_L = (\mu\nabla x^n,\nabla\zeta_h^n)_{L^d} - (\mu\zeta_h^n, n\cdot\nabla x^n)_{L,\partial\Omega} \lesssim \|x^n\|_{A*}\|\zeta_h^n\|_A$),

$$\begin{aligned}\tau(\Psi_h^n,\zeta_h^n)_L &\le \tau\|\tilde\Psi_h^n\|_L\|\zeta_h^n\|_L + \tau\|Bx^n\|_L\|\zeta_h^n\|_L + \tau\|x^n\|_{A*}\|\zeta_h^n\|_A\\ &\le \tau t_*(\|\tilde\Psi_h^n\|_L^2 + \|Bx^n\|_L^2) + \tau t_*^{-1}\|\zeta_h^n\|_L^2 + \tfrac{1}{160}\tau\|\zeta_h^n\|_a^2 + C\tau\|x^n\|_{A*}^2.\end{aligned}$$

Using the bound (42b) on $\|\tilde\Psi^n\|_L + \|Bx^n\|_L + t_*^{-1/2}\|x^n\|_{A*}$, we obtain

$$\tau(\Psi_h^n,\zeta_h^n)_L \le \tfrac{1}{160}\tau\|\zeta_h^n\|_a^2 + \tau t_*^{-1}\|\zeta_h^n\|_L^2 + C\tau(C_\Psi^n)^2\tau^3,$$

so that owing to the bound (54) on ζ_h^n, $\tau \le t_*$, and the definition of E_h^n,

$$\tau(\Psi_h^n,\zeta_h^n)_L \le \tfrac{1}{160}\tau\|\zeta_h^n\|_a^2 + C\tau(E_h^n)^2. \tag{58}$$ eq:bnd.pp

Collecting the bounds (55), (56), (57), and (58) yields the assertion. □

Combining Lemmata 4.1 and 4.2 yields our basic energy estimate.

c.engy **Theorem 4.1** (Basic energy estimate)**.** *Assume* $\gamma \in [\frac{1}{5},\frac{2}{5}]$ *and* $\mathrm{Co} \le 1$. *Then,*

$$\begin{aligned}&\tfrac{1}{2}\|\xi_h^{n+1}\|_L^2 - \tfrac{1}{2}\|\xi_h^n\|_L^2 + \tfrac{1}{2}\|\theta_h^n - \xi_h^n\|_L^2 + \tfrac{3}{8}\tau|\theta_h^n|_S^2 + \tfrac{3}{8}\tau|\zeta_h^n|_S^2 + (\tfrac{1}{2}-\gamma)\tau\|\theta_h^n\|_a^2 + \tfrac{1}{20}\tau\|\zeta_h^n\|_a^2 + \tfrac{1}{40}\tau\|\zeta_h^n+\theta_h^n\|_a^2\\ &\le \tfrac{1}{2}\|\xi_h^{n+1}-\zeta_h^n\|_L^2 + C\tau(E_h^n)^2.\end{aligned} \tag{59}$$ eq:energy.st

Proof. Using the energy identity (46) together with the fact that $\frac{1}{2}-\gamma \ge \frac{1}{10}$ and $\gamma \ge \frac{1}{5}$, and accounting for the bound (52) on the source terms yields

$$\begin{aligned}&\tfrac{1}{2}\|\xi_h^{n+1}\|_L^2 - \tfrac{1}{2}\|\xi_h^n\|_L^2 + \tfrac{1}{2}\|\theta_h^n - \xi_h^n\|_L^2 + \tfrac{3}{8}\tau|\theta_h^n|_S^2 + \tfrac{3}{8}\tau|\zeta_h^n|_S^2 + (\tfrac{1}{2}-\gamma)\tau\|\theta_h^n\|_a^2 + \tfrac{1}{10}\tau\|\zeta_h^n\|_a^2 + \tfrac{1}{10}\tau\|\zeta_h^n+\theta_h^n\|_a^2\\ &\le \tfrac{1}{2}\|\xi_h^{n+1}-\zeta_h^n\|_L^2 + \tfrac{1}{40}\tau\|\theta_h^n\|_a^2 + \tfrac{1}{80}\tau\|\zeta_h^n\|_a^2 + C\tau(E_h^n)^2.\end{aligned}$$

Since the term involving $\|\theta_h^n\|_a^2$ on the left-hand side will be used later in a different context, we leave it as it stands and use instead the terms $\|\zeta_h^n\|_a^2$ and $\|\zeta_h^n+\theta_h^n\|_a^2$ on the left-hand side to absorb the two terms with the $\|\cdot\|_a$-norm on the right-hand side. We observe that

$$\|\theta_h^n\|_a^2 = \|\theta_h^n + \zeta_h^n - \zeta_h^n\|_a^2 \le \tfrac{3}{2}\|\zeta_h^n\|_a^2 + 3\|\zeta_h^n+\theta_h^n\|_a^2$$

to infer the assertion. □

The way to tackle the anti-dissipative term $\frac{1}{2}\|\xi_h^{n+1}-\zeta_h^n\|_L^2$ on the right-hand side of the basic energy estimate (59) depends on the flow regime and will be examined in the two subsequent sections.

ab.adv

4.3. **Stability and convergence: advection-dominated regime**

In this regime, we assume that $\mathrm{Pe} \geq 1$ and, as before to fix the ideas, that $\gamma \in [\frac{1}{5}, \frac{2}{5}]$. Taking a larger interval for γ in $(0, \frac{1}{2})$ is again possible, and this will only modify the numerical factors in the bound on the Courant number. In the advection-dominated regime, an important ingredient to bound the diffusion operator is that there is C_A such that for all $v_h \in V_h$,

$$\tau \|A_h v_h\|_L \leq C_A (\mathrm{Co}/\mathrm{Pe})^{1/2} \tau^{1/2} \|v_h\|_A, \tag{60}$$

eq:C.A

since owing to (27), $\tau \|A_h v_h\|_L \lesssim \tau \mu^{1/2} h^{-1} \|v_h\|_A$ and $\tau^{1/2} \mu^{1/2} h^{-1} = (\mathrm{Co}/\mathrm{Pe})^{1/2}$.

Our first step is to control the anti-dissipative term $\frac{1}{2}\|\xi_h^{n+1} - \zeta_h^n\|_L^2$ on the right-hand side of the basic energy estimate (59). We recall the following inverse inequality valid for piecewise affine functions: There is C_i such that for all $v_h \in V_h$,

$$\|\nabla v_h\|_{L^d} \leq C_i h^{-1} \|v_h - \pi_h^0 v_h\|_L. \tag{61}$$

eq:Cinv

dv.dom **Lemma 4.3** (Stability)**.** *Assume* $\mathrm{Pe} \geq 1$, $\gamma \in [\frac{1}{5}, \frac{2}{5}]$, *and* $\mathrm{Co} \leq 1$. *Assume further that*

$$\mathrm{Co} \leq \min\left\{ \tfrac{1}{2}(C_i C_B)^{-2/3}, \tfrac{1}{8} C_S^{-2}, \tfrac{5}{4} c_a (2C_i + 3)^{-2} C_A^{-2} \mathrm{Pe} \right\}, \tag{62}$$

eq:Co.adv

recalling that C_B *and* C_S *are defined in Lemma* 2.2. *Then,*

$$\tfrac{1}{2}\|\xi_h^{n+1}\|_L^2 - \tfrac{1}{2}\|\xi_h^n\|_L^2 + \tfrac{1}{2}\|\theta_h^n - \xi_h^n\|_L^2 + \tfrac{1}{8}\tau |\theta_h^n|_S^2 + \tfrac{1}{8}\tau |\zeta_h^n|_S^2 + \tfrac{1}{20} c_a \tau \|\theta_h^n\|_A^2 + \tfrac{1}{40} c_a \tau \|\zeta_h^n + \theta_h^n\|_A^2 \lesssim \tau (E_h^n)^2. \tag{63}$$

eq:energy.st

Proof. We start from the basic energy estimate (59) and observe that $\frac{1}{2} - \gamma \geq \frac{1}{10}$ to write

$$\begin{aligned}
&\tfrac{1}{2}\|\xi_h^{n+1}\|_L^2 - \tfrac{1}{2}\|\xi_h^n\|_L^2 + \tfrac{1}{2}\|\theta_h^n - \xi_h^n\|_L^2 + \tfrac{3}{8}\tau |\theta_h^n|_S^2 + \tfrac{3}{8}\tau |\zeta_h^n|_S^2 + \tfrac{1}{10} c_a \tau \|\theta_h^n\|_A^2 + \tfrac{1}{20} c_a \tau \|\zeta_h^n\|_A^2 + \tfrac{1}{40} c_a \tau \|\zeta_h^n + \theta_h^n\|_A^2 \\
&\leq \tfrac{1}{2}\|\xi_h^{n+1} - \zeta_h^n\|_L^2 + C\tau (E_h^n)^2,
\end{aligned}$$

where we have used (18) to replace the $\|\cdot\|_a$-norm by the $\|\cdot\|_A$-norm. Set $\eta_h^n := \theta_h^n - \zeta_h^n$, so that by (16b) and (16c),

$$\xi_h^{n+1} - \zeta_h^n = \tfrac{1}{2}\tau B_h \eta_h^n + \left(\tfrac{1}{2} - 2\gamma\right) \tau A_h \theta_h^n - \left(\tfrac{1}{2} - \gamma\right) \tau A_h \zeta_h^n - \tfrac{1}{2}\tau \beta_h^n + \tau \delta_h^n - \tau \Psi_h^n. \tag{64}$$

eq:xi-zeta

Using the triangle inequality and the bound (21) on B_h yields

$$\begin{aligned}
\|\xi_h^{n+1} - \zeta_h^n\|_L \leq{}& \tfrac{1}{2}\sigma\tau \|\nabla \eta_h^n\|_{L^d} + \tfrac{1}{2} C_S \mathrm{Co}^{1/2} \tau^{1/2} |\eta_h^n|_S + \left|\tfrac{1}{2} - 2\gamma\right| \tau \|A_h \theta_h^n\|_L + \left(\tfrac{1}{2} - \gamma\right) \tau \|A_h \zeta_h^n\|_L \\
&+ \tau (\tfrac{1}{2}\|\beta_h^n\|_L + \|\delta_h^n\|_L + \|\Psi^n\|_L).
\end{aligned}$$

The terms involving the discrete operator A_h are bounded using (60), $\left|\frac{1}{2} - 2\gamma\right| \leq \frac{3}{10}$, and $\left(\frac{1}{2} - \gamma\right) \leq \frac{3}{10}$ yielding

$$\left|\tfrac{1}{2} - 2\gamma\right| \tau \|A_h \theta_h^n\|_L + \left(\tfrac{1}{2} - \gamma\right) \tau \|A_h \zeta_h^n\|_L \leq \tfrac{3}{10} \tau^{1/2} C_A (\mathrm{Co}/\mathrm{Pe})^{1/2} (\|\theta_h^n\|_A + \|\zeta_h^n\|_A).$$

The contributions of A_h to β_h^n and δ_h^n are bounded using (26) and $\tau^{1/2} \mu^{1/2} h^{-1} = (\mathrm{Co}/\mathrm{Pe})^{1/2} \leq 1$ so that

$$\tau \|A_h \theta_\pi^n\|_L + \tau \|A_h \zeta_\pi^n\|_L \lesssim \tau \mu^{1/2} h^{-1} (\|\theta_\pi^n\|_{A*} + \|\zeta_\pi^n\|_{A*}) \leq \tau^{1/2} E_h^n.$$

The contributions of B_h to β_h^n and δ_h^n are bounded using (25) and $\mathrm{Co} \leq 1$ so that

$$\tau \|B_h \theta_\pi^n\|_L + \tau \|B_h \zeta_\pi^n\|_L \lesssim \tau^{1/2} (\|\theta_\pi^n\|_{B*} + \|\zeta_\pi^n\|_{B*}) \leq \tau^{1/2} E_h^n.$$

Hence,

$$\tau \|\beta_h^n\|_L + \tau \|\delta_h^n\|_L \lesssim \tau^{1/2} E_h^n. \tag{65}$$

eq:bnd.beta.

Finally,

$$\tau\|\Psi_h^n\|_L \le \tau\|\Psi^n\|_L \lesssim \tau\tilde{C}_\Psi^n\tau \le \tau^{1/2}C_\Psi^n\tau^{3/2} \le \tau^{1/2}E_h^n, \tag{66}$$

eq:bnd.Psi.b

owing to the bound (42a) on $\|\Psi^n\|_L$ and the fact that $\tilde{C}_\Psi^n \le C_\Psi^n$. As a result,

$$\|\xi_h^{n+1}-\zeta_h^n\|_L \le \tfrac{1}{2}\sigma\tau\|\nabla\eta_h^n\|_{L^d} + \tfrac{1}{2}C_S\mathrm{Co}^{1/2}\tau^{1/2}|\eta_h^n|_S + \tfrac{3}{10}C_A(\mathrm{Co/Pe})^{1/2}\tau^{1/2}(\|\theta_h^n\|_A+\|\zeta_h^n\|_A) + C\tau^{1/2}E_h^n. \tag{67}$$

eq:step1

The next step is to control $\|\nabla\eta_h^n\|_{L^d}$. Let $\varsigma_h^n = \eta_h^n - \pi_h^0\eta_h^n$ and observe that

$$\|\varsigma_h^n\|_L^2 = (\eta_h^n,\varsigma_h^n)_L = \tau(B_h\theta_h^n,\varsigma_h^n)_L + (1-3\gamma)\tau(A_h\theta_h^n,\varsigma_h^n)_L + \gamma\tau(A_h\zeta_h^n,\varsigma_h^n)_L - \tau(\beta_h^n,\varsigma_h^n)_L,$$

since $\eta_h^n = \tau B_h\theta_h^n + (1-3\gamma)\tau A_h\theta_h^n + \gamma\tau A_h\zeta_h^n - \tau\beta_h^n$ owing to (16b). To bound the first term on the right-hand side, we use the bound (23) on B_h to infer

$$\tau|(B_h\theta_h^n,\varsigma_h^n)_L| \le C_B\mathrm{Co}^{1/2}\tau^{1/2}(|\theta_h^n|_S + \sigma_1^{1/2}\|\theta_h^n\|_L)\|\varsigma_h^n\|_L.$$

Furthermore, bounding the three other terms by the Cauchy–Schwarz inequality, using the fact that $\gamma\in[\frac{1}{5},\frac{2}{5}]$ and the bound (60) for the terms involving the discrete operator A_h, and simplifying by $\|\varsigma_h^n\|_L$,

$$\|\varsigma_h^n\|_L \le C_B\mathrm{Co}^{1/2}\tau^{1/2}(|\theta_h^n|_S + \sigma_1^{1/2}\|\theta_h^n\|_L) + \tfrac{2}{5}C_A(\mathrm{Co/Pe})^{1/2}\tau^{1/2}(\|\theta_h^n\|_A+\|\zeta_h^n\|_A) + \tau\|\beta_h^n\|_L,$$

so that using the bound (53) on $\|\theta_h^n\|_L$, $\tau\sigma_1\le 1$, and (65) to bound $\tau\|\beta_h^n\|_L$,

$$\|\varsigma_h^n\|_L \le C_B\mathrm{Co}^{1/2}\tau^{1/2}|\theta_h^n|_S + \tfrac{2}{5}C_A(\mathrm{Co/Pe})^{1/2}\tau^{1/2}(\|\theta_h^n\|_A+\|\zeta_h^n\|_A) + C\tau^{1/2}E_h^n.$$

Thus, using the inverse inequality (61),

$$\begin{aligned}\sigma\tau\|\nabla\eta_h^n\|_{L^d} &\le C_i\sigma\tau h^{-1}\|\varsigma_h^n\|_L = C_i\mathrm{Co}\|\varsigma_h^n\|_L\\ &\le C_iC_B\mathrm{Co}^{3/2}\tau^{1/2}|\theta_h^n|_S + \tfrac{2}{5}C_iC_A\mathrm{Co}(\mathrm{Co/Pe})^{1/2}\tau^{1/2}(\|\theta_h^n\|_A+\|\zeta_h^n\|_A) + C\tau^{1/2}E_h^n.\end{aligned}$$

Substituting back into (67), re-arranging terms, and since $\mathrm{Co}\le 1$, we infer

$$\begin{aligned}\|\xi_h^{n+1}-\zeta_h^n\|_L \le{}& \tfrac{1}{2}C_iC_B\mathrm{Co}^{3/2}\tau^{1/2}|\theta_h^n|_S + \tfrac{1}{2}C_S\mathrm{Co}^{1/2}\tau^{1/2}|\theta_h^n-\zeta_h^n|_S\\ &+ (\tfrac{1}{5}C_i+\tfrac{3}{10})C_A(\mathrm{Co/Pe})^{1/2}\tau^{1/2}(\|\theta_h^n\|_A+\|\zeta_h^n\|_A) + C\tau^{1/2}E_h^n.\end{aligned}$$

Let $\chi_1 := 32^{-1/2}$ and $\chi_2 := 80^{-1/2}$. Then, owing to the assumption (62) on the Courant number, the above inequality becomes

$$\|\xi_h^{n+1}-\zeta_h^n\|_L \le \chi_1\tau^{1/2}(|\theta_h^n|_S + |\theta_h^n-\zeta_h^n|_S) + \chi_2c_a^{1/2}\tau^{1/2}(\|\theta_h^n\|_A+\|\zeta_h^n\|_A) + C\tau^{1/2}E_h^n.$$

Since $|\theta_h^n|_S+|\theta_h^n-\zeta_h^n|_S \le 2(|\theta_h^n|_S+|\zeta_h^n|_S)$, squaring the above bound, and using that $\frac{1}{2}(a+b+c)^2 \le a^2+2b^2+2c^2$ where a, b, and c denote the three addends on the right-hand side of the above equation yields

$$\tfrac{1}{2}\|\xi_h^{n+1}-\zeta_h^n\|_L^2 \le 8\chi_1^2\tau(|\theta_h^n|_S^2+|\zeta_h^n|_S^2) + 4\chi_2^2c_a\tau(\|\theta_h^n\|_A^2+\|\zeta_h^n\|_A^2) + C\tau E_h^n.$$

Finally, observing that $8\chi_1^2 = \frac{1}{4}$ and $4\chi_2^2 = \frac{1}{20}$ yields the assertion. □

Remark 4.2 (Purely advective case). In the purely advective case ($\mu = 0$), the third argument in the bound (62) on the Courant number can be dropped, leading to the bound derived in [8].

Remark 4.3 (Parabolic CFL restriction). In the advection-dominated regime, there holds $\tau\mu h^{-2} = \mathrm{CoPe}^{-1} \le 1$, that is, a parabolic CFL restriction on the time step. In particular, this property has been used in the proof of Lemma 4.3 to control the terms with the discrete operator A_h using (60). We stress that this property is not used in the diffusion-dominated regime, where it will be too restrictive.

We can now derive our main convergence result in the advection-dominated regime.

dv.dom **Theorem 4.2** (Convergence in L-norm)**.** *With the basic assumptions stated in Section* 2.1*, assume* $\mathrm{Pe} \ge 1$*, take* $\gamma \in [\frac{1}{5}, \frac{2}{5}]$*, and assume the bound* (62) *on the Courant number. Then,*

$$\|u^N - u_h^N\|_L \lesssim C_{\mathrm{tim}}\tau^{3/2} + C_{\mathrm{spc}}\sigma^{1/2}h^{3/2}, \tag{68}$$ eq:err.est.a

where $C_{\mathrm{tim}}^2 = \sum_{n=0}^{N-1} \tau (C_\Psi^n)^2$ *with* C_Ψ^n *defined by* (41a) *and* $C_{\mathrm{spc}}^2 = \sum_{n=0}^{N-1} \tau((\tilde K_2^n)^2 + (\sigma^{-1}K_{w-u}^n)^2)$ *with* $\tilde K_2^n$ *and* K_{w-u}^n *defined in Lemmata* 3.2 *and* 3.3 *respectively.*

Proof. Using the stability result of Lemma 4.3, we sum over n, discard the dissipative terms on the left-hand side, and use a discrete Gronwall lemma to eliminate the contribution of $\|\xi_h^n\|_L^2$ in E_h^n. This yields

$$\|\xi_h^N\|_L^2 \lesssim \sum_{n=0}^{N-1} \tau(\|\theta_\pi^n\|_{B*}^2 + \|\zeta_\pi^n\|_{B*}^2 + \|\theta_\pi^n\|_{A*}^2 + \|\zeta_\pi^n\|_{A*}^2 + (C_\Psi^n)^2\tau^3).$$

To bound the terms with θ_π^n and ζ_π^n, we use the result of Lemma 3.7, and the fact that $\mu^{1/2} \le \sigma^{1/2}h^{1/2}$ since $\mathrm{Pe} \ge 1$ and $\tau^{1/2}hK_{w-u}^n \le \sigma^{1/2}h^{3/2}(\sigma^{-1}K_{w-u}^n)$ since $\mathrm{Co} \le 1$. This yields $\|\xi_h^N\|_L \lesssim C_{\mathrm{tim}}\tau^{3/2} + C_{\mathrm{spc}}\sigma^{1/2}h^{3/2}$ and we conclude using the triangle inequality. □

The convergence result of Theorem 4.2 can be completed by showing additionally convergence in the $\|\cdot\|_A$-norm. The proof is postponed to §7.1.

dv.dom **Proposition 4.1** (Convergence in $\|\cdot\|_A$-norm)**.** *Under the assumptions of Theorem* 4.2*, there holds*

$$\left(\tau\sum_{n=1}^{N}\|u^n - u_h^n\|_A^2\right)^{1/2} \lesssim C_{\mathrm{tim}}\tau^{3/2} + C_{\mathrm{spc}}\sigma^{1/2}h^{3/2}.$$

4.4. **Stability and convergence: diffusion-dominated regime**

ab.dif In this regime, we assume $\mathrm{Pe} \le 1$. We derive three intermediate stability results. First (Lemma 4.4), we tighten the basic energy estimate (59) by achieving additional control on the increment $\|\theta_h^n - \zeta_h^n\|_L^2$. Then (Lemma 4.6), we bound the anti-dissipative term $\frac{1}{2}\|\xi_h^{n+1} - \zeta_h^n\|_L^2$. Finally (Lemma 4.7), we achieve additional control on $\tau\|\xi_h^{n+1}\|_A^2$. For our first step, it is sufficient that $\gamma \in (\frac{1}{4}, \frac{2}{5}]$; the minimal threshold on γ serves to obtain only positive factors on the left-hand side of the new energy estimate (70). For our second and third steps, we need the parameter γ to be sufficiently close to $\gamma_* = 1 - \frac{1}{\sqrt{2}} \simeq 0.293$. For simplicity, we assume $\gamma = \gamma_*$ and postpone to Remark 4.5 the discussion when γ slightly deviates from γ_*, as motivated for instance by finite arithmetic precision.

In the diffusion-dominated regime, an important ingredient to bound the operator B_h is that there is C_{BA} such that for all $v_h \in V_h$,

$$\tau\|B_hv_h\|_L \le C_{BA}\tau\sigma\mu^{-1/2}\|v_h\|_A = C_{BA}(\mathrm{CoPe})^{1/2}\tau^{1/2}\|v_h\|_A, \tag{69}$$ eq:C.BA

since owing to the definition of the $\|\cdot\|_A$-norm, $\|\nabla v_h\|_{L^d} \le \mu^{-1/2}\|v_h\|_A$, while a discrete trace inequality yields $|v_h|_S \lesssim (\frac{\sigma h}{\mu})^{1/2}\|v_h\|_A$, so that (69) results from (21) and $\tau^{1/2}\sigma\mu^{-1/2} = (\mathrm{CoPe})^{1/2}$.

ergy.+

Lemma 4.4. *Assume* $\gamma \in (\frac{1}{4}, \frac{2}{5}]$. *Assume* $\mathrm{Co} \leq \min(1, \frac{1}{30} c_a C_{BA}^{-2} \mathrm{Pe}^{-1})$. *Then,*

$$\begin{aligned} &\tfrac{1}{2}\|\xi_h^{n+1}\|_L^2 - \tfrac{1}{2}\|\xi_h^n\|_L^2 + \tfrac{1}{2}\|\theta_h^n - \xi_h^n\|_L^2 + \tfrac{3}{8}\|\theta_h^n - \zeta_h^n\|_L^2 + \tfrac{1}{8}\tau|\theta_h^n|_S^2 + \tfrac{1}{8}\tau|\zeta_h^n|_S^2 + \tfrac{1}{8}(\tfrac{1}{2} - \gamma)c_a\tau\|\theta_h^n\|_A^2 \\ &+ \tfrac{1}{8}c_a\tau\|\zeta_h^n\|_A^2 + \tfrac{3}{4}(\gamma - \tfrac{1}{4})c_a\tau\|\theta_h^n - \zeta_h^n\|_A^2 + \tfrac{1}{40}c_a\tau\|\zeta_h^n + \theta_h^n\|_A^2 \leq \tfrac{1}{2}\|\xi_h^{n+1} - \zeta_h^n\|_L^2 + C\tau(E_h^n)^2. \end{aligned} \tag{70}$$

energy.+

Proof. We first observe that (16b) implies

$$\theta_h^n - \zeta_h^n = \tau B_h\theta_h^n + (1 - 3\gamma)\tau A_h\theta_h^n + \gamma\tau A_h\zeta_h^n - \tau\beta_h^n,$$

and re-arranging terms leads to

$$\theta_h^n - \zeta_h^n = \tau B_h\theta_h^n - (\gamma - \tfrac{1}{2})\tau A_h(\theta_h^n + \zeta_h^n) - (2\gamma - \tfrac{1}{2})\tau A_h(\theta_h^n - \zeta_h^n) - \tau\beta_h^n.$$

Taking the L-scalar product with $\theta_h^n - \zeta_h^n$ and using the symmetry of a_h yields

$$\|\theta_h^n - \zeta_h^n\|_L^2 = \tau(B_h\theta_h^n, \theta_h^n - \zeta_h^n)_L - (\gamma - \tfrac{1}{2})\tau(\|\theta_h^n\|_a^2 - \|\zeta_h^n\|_a^2) - (2\gamma - \tfrac{1}{2})\tau\|\theta_h^n - \zeta_h^n\|_a^2 - \tau(\beta_h^n, \theta_h^n - \zeta_h^n)_L.$$

Since $\tau(B_h\theta_h^n, \theta_h^n - \zeta_h^n)_L \leq \frac{1}{2}\tau^2\|B_h\theta_h^n\|_L^2 + \frac{1}{2}\|\theta_h^n - \zeta_h^n\|_L^2$, this yields, re-arranging terms,

$$\tfrac{1}{2}\|\theta_h^n - \zeta_h^n\|_L^2 + (\gamma - \tfrac{1}{2})\tau(\|\theta_h^n\|_a^2 - \|\zeta_h^n\|_a^2) + (2\gamma - \tfrac{1}{2})\tau\|\theta_h^n - \zeta_h^n\|_a^2 \leq \tfrac{1}{2}\tau^2\|B_h\theta_h^n\|_L^2 - \tau(\beta_h^n, \theta_h^n - \zeta_h^n)_L. \tag{71}$$

add.energy

The idea is now to combine (71) with (59) so as to absorb the positive term $\frac{1}{2}\tau^2\|B_h\theta_h^n\|_L^2$ by dissipative terms on the left-hand side. To this purpose, we multiply (71) by a real number $\alpha \in (0, 1)$ and sum the resulting estimate to (59). To fix the ideas, we take $\alpha = \frac{3}{4}$ yielding

$$\begin{aligned} &\tfrac{1}{2}\|\xi_h^{n+1}\|_L^2 - \tfrac{1}{2}\|\xi_h^n\|_L^2 + \tfrac{1}{2}\|\theta_h^n - \xi_h^n\|_L^2 + \tfrac{3}{8}\|\theta_h^n - \zeta_h^n\|_L^2 + \tfrac{3}{8}\tau|\theta_h^n|_S^2 + \tfrac{3}{8}\tau|\zeta_h^n|_S^2 \\ &+ \tfrac{1}{4}(\tfrac{1}{2} - \gamma)c_a\tau\|\theta_h^n\|_A^2 + \tfrac{1}{8}c_a\tau\|\zeta_h^n\|_A^2 + \tfrac{3}{4}(2\gamma - \tfrac{1}{2})c_a\tau\|\theta_h^n - \zeta_h^n\|_A^2 + \tfrac{1}{40}c_a\tau\|\zeta_h^n + \theta_h^n\|_A^2 \\ &\leq \tfrac{1}{2}\|\xi_h^{n+1} - \zeta_h^n\|_L^2 + \tfrac{3}{8}\tau^2\|B_h\theta_h^n\|_L^2 - \tfrac{3}{4}\tau(\beta_h^n, \theta_h^n - \zeta_h^n)_L + C\tau(E_h^n)^2, \end{aligned}$$

where we have used the discrete stability (18) of A_h to substitute the $\|\cdot\|_a$-norms by $\|\cdot\|_A$-norms on the left-hand side and the fact that $\frac{1}{2} - \gamma \geq \frac{1}{10}$ to simplify the term with $\|\zeta_h^n\|_A^2$. Using (69) and the assumption on the Courant number yields

$$\tfrac{3}{8}\tau^2\|B_h\theta_h^n\|_L^2 \leq \tfrac{1}{80}c_a\tau\|\theta_h^n\|_A^2 = \tfrac{1}{8}\left\{\min_{\gamma\in(\frac{1}{4},\frac{2}{5}]}(\tfrac{1}{2} - \gamma)\right\}c_a\tau\|\theta_h^n\|_A^2,$$

so that this term can be absorbed using half of the $\|\theta_h^n\|_A^2$-term on the left-hand side of the above energy estimate. Finally, we bound $\frac{3}{4}\tau(\beta_h^n, \theta_h^n - \zeta_h^n)_L$. We obtain using the boundedness (22) and (26) of B_h and A_h,

$$\tfrac{3}{4}\tau|(\beta_h^n, \theta_h^n - \zeta_h^n)_L| \lesssim \tau\|\theta_\pi^n\|_{B*}(|\theta_h^n - \zeta_h^n|_S + \sigma_1^{1/2}\|\theta_h^n - \zeta_h^n\|_L) + \tau(\|\theta_\pi^n\|_{A*} + \|\zeta_\pi^n\|_{A*})\|\theta_h^n - \zeta_h^n\|_A.$$

The first term is bounded as

$$\begin{aligned} \tau\|\theta_\pi^n\|_{B*}(|\theta_h^n - \zeta_h^n|_S + \sigma_1^{1/2}\|\theta_h^n - \zeta_h^n\|_L) &\leq \tfrac{1}{4}\tau(|\theta_h^n|_S^2 + |\zeta_h^n|_S^2) + C\tau(\|\theta_\pi^n\|_{B*}^2 + \sigma_1\|\theta_h^n\|_L^2 + \sigma_1\|\zeta_h^n\|_L^2) \\ &\leq \tfrac{1}{4}\tau(|\theta_h^n|_S^2 + |\zeta_h^n|_S^2) + C\tau(E_h^n)^2, \end{aligned}$$

where we have used $\tau\sigma_1 \leq 1$ and the bounds (53) and (54) on $\|\theta_h^n\|_L$ and $\|\zeta_h^n\|_L$. For the second term,

$$\tau(\|\theta_\pi^n\|_{A*} + \|\zeta_\pi^n\|_{A*})\|\theta_h^n - \zeta_h^n\|_A \leq \tfrac{3}{4}(\gamma - \tfrac{1}{4})\tau\|\theta_h^n - \zeta_h^n\|_A^2 + C\tau(\|\theta_\pi^n\|_{A*}^2 + \|\zeta_\pi^n\|_{A*}^2).$$

Collecting the above estimates yields the assertion. □

Our second step aims at controlling the anti-dissipative term $\frac{1}{2}\|\xi_h^{n+1}-\zeta_h^n\|_L^2$ on the right-hand side of the energy estimate (70). To this purpose, it is useful to reformulate the last step (16c) of the error equation without using the discrete operator A_h. We simply state the result, since the proof amounts to a direct verification.

reform **Lemma 4.5** (Reformulation of last step without A_h)**.** *Let* $\omega_1 := \gamma^{-1}(\frac{1}{2}-\gamma)$ *and* $\omega_2 := \frac{1}{2\gamma^2}(-1+4\gamma-2\gamma^2)$. *Then,*

$$\xi_h^{n+1}-\zeta_h^n = \omega_1(\zeta_h^n-\theta_h^n)+\omega_2(\theta_h^n-\xi_h^n)-\tfrac{1}{2}\tau B_h(\zeta_h^n-\theta_h^n)+\omega_1\tau B_h\theta_h^n-\tau\Xi_h^n-\tau\Psi_h^n, \tag{72}$$

inc_diff

where

$$\Xi_h^n := -\tfrac{1}{2}B_h(\zeta_\pi^n-\theta_\pi^n)+\omega_1 B_h\theta_\pi^n. \tag{73}$$

eq:def.Xi

In what follows, we assume $\gamma=\gamma_*$. An important fact used hereafter is that $\omega_2(\gamma_*)=0$, thereby zeroing out the contribution of ξ_h^n on the right-hand side of (72). We are now ready to bound the anti-dissipative term. Note that we tighten the assumption on the Courant number with respect to Lemma 4.4.

rgy.++ **Lemma 4.6.** *Assume* $\gamma=\gamma_*$ *and*

$$\mathrm{Co}\le\min(1,\tfrac{1}{180}c_aC_{BA}^{-2}\mathrm{Pe}^{-1}). \tag{74}$$

eq:Co.dif

Then,

$$\tfrac{1}{2}\|\xi_h^{n+1}\|_L^2-\tfrac{1}{2}\|\xi_h^n\|_L^2+\tfrac{1}{8}\tau|\theta_h^n|_S^2+\tfrac{1}{8}\tau|\zeta_h^n|_S^2+\tfrac{1}{80}c_a\tau\|\theta_h^n\|_A^2+\tfrac{1}{8}c_a\tau\|\zeta_h^n\|_A^2+\tfrac{1}{40}c_a\tau\|\zeta_h^n+\theta_h^n\|_A^2\lesssim\tau(E_h^n)^2. \tag{75}$$

energy.++

Proof. We start from the result of Lemma 4.5. Observing that $\omega_1=\frac{1}{\sqrt{2}}$ and $\omega_2=0$ for $\gamma=\gamma_*$ and setting

$$X_h^n = -\tfrac{1}{2}B_h(\zeta_h^n-\theta_h^n)+\tfrac{1}{\sqrt{2}}B_h\theta_h^n-\Xi_h^n-\Psi_h^n,$$

where Ξ_h^n is defined by (73), we infer

$$\xi_h^{n+1}-\zeta_h^n=\frac{1}{\sqrt{2}}(\zeta_h^n-\theta_h^n)+\tau X_h^n.$$

This yields for positive real number ϵ, $\frac{1}{2}\|\xi_h^{n+1}-\zeta_h^n\|_L^2\le\frac{1}{2}(1+\epsilon^{-1})\tau^2\|X_h^n\|_L^2+\frac{1}{4}(1+\epsilon)\|\zeta_h^n-\theta_h^n\|_L^2$. Choosing $\epsilon=\frac{1}{2}$, we infer

$$\tfrac{1}{2}\|\xi_h^{n+1}-\zeta_h^n\|_L^2\le\tfrac{3}{2}\tau^2\|X_h^n\|_L^2+\tfrac{3}{8}\|\zeta_h^n-\theta_h^n\|_L^2.$$

We now bound the term $\|X_h^n\|_L^2$. Since $\frac{1}{3}(a+b+c)^2\le a^2+b^2+c^2$ for real numbers a, b, and c, we obtain using (69),

$$\begin{aligned}\tfrac{1}{3}\tau^2\|X_h^n\|_L^2&\le\tfrac{1}{4}\tau^2\|B_h(\zeta_h^n-\theta_h^n)\|_L^2+\tfrac{1}{2}\tau^2\|B_h\theta_h^n\|_L^2+\tau^2\|\Xi_h^n+\Psi_h^n\|_L^2\\&\le C_{BA}^2(\mathrm{CoPe})\tau(\tfrac{1}{4}\|\zeta_h^n-\theta_h^n\|_A^2+\tfrac{1}{2}\|\theta_h^n\|_A^2)+\tau^2\|\Xi_h^n+\Psi_h^n\|_L^2.\end{aligned}$$

Owing to (25) and $\mathrm{Co}\le1$, $\tau\|\Xi_h^n\|_L\lesssim\tau^{1/2}E_h^n$ and recalling $\tau\|\Psi_h^n\|_L\lesssim\tau^{1/2}E_h^n$ from (66), we obtain

$$\tfrac{1}{3}\tau^2\|X_h^n\|_L^2\le C_{BA}^2(\mathrm{CoPe})\tau(\tfrac{1}{4}\|\zeta_h^n-\theta_h^n\|_A^2+\tfrac{1}{2}\|\theta_h^n\|_A^2)+C\tau(E_h^n)^2.$$

Owing to the assumption on the Courant number,

$$3\tfrac{3}{2}\tfrac{1}{4}C_{BA}^2c_a^{-1}(\mathrm{CoPe})\le\tfrac{1}{160}\le\tfrac{3}{4}(\gamma_*-\tfrac{1}{4}),\qquad 3\tfrac{3}{2}\tfrac{1}{2}C_{BA}^2c_a^{-1}(\mathrm{CoPe})\le\tfrac{1}{80}\le\tfrac{1}{16}(\tfrac{1}{2}-\gamma_*).$$

As a result,

$$\tfrac{1}{2}\|\xi_h^{n+1}-\zeta_h^n\|_L^2\le\tfrac{3}{4}(\gamma_*-\tfrac{1}{4})c_a\tau\|\zeta_h^n-\theta_h^n\|_A^2+\tfrac{1}{16}(\tfrac{1}{2}-\gamma_*)c_a\tau\|\theta_h^n\|_A^2+\tfrac{3}{8}\|\zeta_h^n-\theta_h^n\|_L^2+C\tau(E_h^n)^2.$$

Using this estimate in (70) yields the assertion since $\frac{1}{16}(\frac{1}{2}-\gamma_*)\ge\frac{1}{80}$. □

We can now proceed to our third and final step in the stability analysis. Our goal is to infer a control on $\tau\|\xi_h^{n+1}\|_A^2$ from the control on $\tau\|\theta_h^n\|_A^2$ and $\tau\|\zeta_h^n\|_A^2$ achieved in (75). This will require replacing the quantity E_h^n by

$$\tilde{E}_h^n := t_*^{-1/2}\|\xi_h^n\|_L + \|\theta_\pi^n\|_{B*} + \|\theta_\pi^n\|_{A*} + \|\zeta_\pi^n\|_{B*} + \|\zeta_\pi^n\|_{A*} + \mathrm{Pe}^{-1/2}(|\theta_\pi^n|_S + |\zeta_\pi^n|_S) + t_*^{1/2}\tilde{C}_\Psi^n\tau. \tag{76}$$

eq:tilde.E

The definition of $\tilde{E}_h^n$ entails two modifications with respect to E_h^n. Firstly, the term $\mathrm{Pe}^{-1/2}(|\theta_\pi^n|_S + |\zeta_\pi^n|_S)$ has been added; this change will not modify the convergence rate in space with respect to the $\|\cdot\|_{A*}$- and $\|\cdot\|_{B*}$-norms of θ_π^n and ζ_π^n. Secondly, and more importantly, the time error is now of lower-order since the term $C_\Psi^n\tau^{3/2}$ has been replaced by $t_*^{1/2}\tilde{C}_\Psi^n\tau$.

gy.+++

Lemma 4.7. *Assume $\gamma = \gamma_*$ and the bound (74) on the Courant number. Assume the additional hyperbolic-type restriction on the time step,*

$$\tau \le t_*^{1/2}\mu^{-1/2}h. \tag{77}$$

eq:CFL.mu

Then,

$$\tfrac{1}{2}\|\xi_h^{n+1}\|_L^2 - \tfrac{1}{2}\|\xi_h^n\|_L^2 + \tfrac{1}{8}\tau|\theta_h^n|_S^2 + \tfrac{1}{8}\tau|\zeta_h^n|_S^2 + \tfrac{1}{80}c_a\tau\|\xi_h^{n+1}\|_A^2 \lesssim \tau(\tilde{E}_h^n)^2. \tag{78}$$

energy.+++

Proof. We take the L-scalar product of (72) with $\tau A_h\xi_h^{n+1}$ to infer

$$\tau(\xi_h^{n+1}, A_h\xi_h^{n+1})_L = \tau(T_1 + T_2, A_h\xi_h^{n+1})_L + \tau^2(T_3 - \Xi_h^n - \Psi_h^n, A_h\xi_h^{n+1})_L, \tag{79}$$

eq:xi.A.xi

where

$$T_1 = (1+\omega_1)\zeta_h^n + (\omega_2 - \omega_1)\theta_h^n, \qquad T_2 = -\omega_2\xi_h^n, \qquad T_3 = -\tfrac{1}{2}B_h(\zeta_h^n - \theta_h^n) + \omega_1 B_h\theta_h^n.$$

Since $\gamma = \gamma_*$, $\omega_2 = 0$ so that $T_2 = 0$. We now bound the other terms on the right-hand side of (79). To bound the term with T_1, we use the boundedness (27) of A_h to infer

$$\tau(T_1, A_h\xi_h^{n+1})_L \lesssim \tau(\|\theta_h^n\|_A + \|\zeta_h^n\|_A)\|\xi_h^{n+1}\|_A.$$

To bound the term with T_3, we use the Cauchy–Schwarz inequality, (69), (27), and $\mathrm{Co} \le 1$, yielding

$$\tau^2(B_h\theta_h^n, A_h\xi_h^{n+1})_L \le \tau^2\|B_h\theta_h^n\|_L\|A_h\xi_h^{n+1}\|_L \lesssim \tau^2\sigma\mu^{-1/2}\|\theta_h^n\|_A\mu^{1/2}h^{-1}\|\xi_h^{n+1}\|_A \le \tau\|\theta_h^n\|_A\|\xi_h^{n+1}\|_A.$$

Proceeding similarly for the contribution of ζ_h^n, we infer

$$\tau^2(T_3, A_h\xi_h^{n+1})_L \lesssim \tau(\|\theta_h^n\|_A + \|\zeta_h^n\|_A)\|\xi_h^{n+1}\|_A.$$

To bound the term with Ξ_h^n, recalling (73), we first observe using (27) that

$$\tau^2(B_h\theta_\pi^n, A_h\xi_h^{n+1})_L \lesssim \tau^2\|B_h\theta_\pi^n\|_A\|\xi_h^{n+1}\|_A \lesssim \tau(\|\theta_\pi^n\|_{A*} + (\tfrac{\mu}{\sigma h})^{1/2}|\theta_\pi^n|_S)\|\xi_h^{n+1}\|_A,$$

since owing to (27), (21), and $\mathrm{Co} \le 1$,

$$\begin{aligned}\tau\|B_h\theta_\pi^n\|_A &\lesssim \tau\mu^{1/2}h^{-1}\|B_h\theta_\pi^n\|_L \lesssim \tau\mu^{1/2}h^{-1}\left(\sigma\|\nabla\theta_\pi^n\|_{L^d} + \sigma^{1/2}h^{-1/2}|\theta_\pi^n|_S\right)\\ &\le \mu^{1/2}\|\nabla\theta_\pi^n\|_{L^d} + (\tfrac{\mu}{\sigma h})^{1/2}|\theta_\pi^n|_S.\end{aligned}$$

Proceeding similarly for the contribution of ζ_π^n, we infer

$$\tau^2(\Xi_h^n, A_h\xi_h^{n+1})_L \lesssim \tau(\|\theta_\pi^n\|_{A*} + \|\zeta_\pi^n\|_{A*} + (\tfrac{\mu}{\sigma h})^{1/2}(|\theta_\pi^n|_S + |\zeta_\pi^n|_S))\|\xi_h^{n+1}\|_A.$$

Finally, to bound the term with Ψ_h^n, we use the Cauchy–Schwarz inequality and (27) to infer

$$\tau^2(\Psi_h^n, A_h\xi_h^{n+1})_L \le \tau^2\|\Psi_h^n\|_L\mu^{1/2}h^{-1}\|\xi_h^{n+1}\|_A \le \tau t_*^{1/2}\|\Psi_h^n\|_L\|\xi_h^{n+1}\|_A \le \tau\tilde{E}_h^n\|\xi_h^{n+1}\|_A,$$

owing to the assumption (77) on the time step and the fact that $t_*^{1/2}\|\Psi_h^n\|_L \le \tilde{E}_h^n$ owing to (42a). Combining the above bounds and using the discrete stability (18) we obtain

$$\tau c_a\|\xi_h^{n+1}\|_A^2 \lesssim \tau\left(\|\theta_h^n\|_A + \|\zeta_h^n\|_A + \tilde{E}_h^n\right)\|\xi_h^{n+1}\|_A,$$

whence the conclusion is straightforward using the stability estimate (75). □

Remark 4.4 (Restrictions on the time step). When the Péclet number is sufficiently large, the condition (74) simply reduces to $\mathrm{Co} \le 1$. In the pure-diffusion limit, this condition, in turn, becomes trivial, and the only restriction on the time step is (77), which is needed to handle the truncation error in Lemma 4.7. Note also that the conditions $\mathrm{Co} \le 1$ and (77) can be regrouped into the condition $\tau \le t_*^{1/2}\bar{\mu}^{-1/2}h$ with $\bar{\mu}$ defined in §2.1.

:gamma

Remark 4.5 (Choice of γ). The parameter γ can slightly deviate from the value γ_*, but this leads to a more stringent bound on the Courant number than (74). Using (72), for positive real numbers ϵ and $\hat{\epsilon}$, we obtain

$$\begin{aligned}\tfrac{1}{2}\|\xi_h^{n+1} - \zeta_h^n\|_L^2 &\le \tfrac{1}{2}(1+\epsilon^{-1})\tau^2\|X_h^n\|_L^2 + \tfrac{1}{2}(1+\epsilon)\|\omega_1(\zeta_h^n - \theta_h^n) + \omega_2(\theta_h^n - \xi_h^n)\|_L^2\\ &\le \tfrac{1}{2}(1+\epsilon^{-1})\tau^2\|X_h^n\|_L^2 + \tfrac{1}{2}(1+\epsilon)(1+\hat{\epsilon})\omega_1^2\|\zeta_h^n - \theta_h^n\|_L^2 + \tfrac{1}{2}(1+\epsilon)(1+\hat{\epsilon}^{-1})\omega_2^2\|\theta_h^n - \xi_h^n\|_L^2.\end{aligned}$$

For $\gamma \in [\frac{1}{4}, \frac{1}{2}]$, ω_1 is a decreasing function of γ taking values in $[0,1]$, while ω_2 is an increasing function of γ taking values in $[-1,1]$ with $\omega_2(\gamma_*) = 0$. The proof of Lemma 4.6 can be extended as long as there is $\hat{\epsilon} > 0$ such that $\frac{1}{2}(1+\hat{\epsilon})\omega_1^2 \le \frac{3}{8}$ and $\frac{1}{2}(1+\hat{\epsilon}^{-1})\omega_2^2 \le \frac{1}{2}$ exploiting the presence of the term $\frac{1}{2}\|\theta_h^n - \zeta_h^n\|_L^2$ on the left-hand side of (70). A direct verification shows that this is possible as long as $\gamma \in (\gamma_{**}, \frac{1}{2})$ with $\gamma_{**} = (2+\sqrt{8/3})^{-1} \simeq 0.275$ (corresponding to $\omega_1 = \sqrt{2/3}$ and $\omega_2 = -1/3$). We observe that the above numerical values depend on the choice $\alpha = \frac{3}{4}$ made in the proof of Lemma 4.4. Taking a larger value for $\alpha < 1$ yields a more stringent bound on the Courant number in Lemma 4.4 but more flexibility in the choice of γ. Finally, the result of Lemma 4.7 is slightly modified since bounding the term $\omega_2\tau(\xi_h^n, A_h\xi_h^{n+1})_L$ by Young's inequality leads to an additional term on the right-hand of (78) of the form $\frac{1}{80}\lambda c_a\tau\|\xi_h^n\|_A^2$ where λ can be chosen < 1 provided γ is sufficiently close to γ_* so that ω_2 is sufficiently small. Details are skipped for brevity.

We can now derive our main convergence result in the diffusion-dominated regime.

if.dom

Theorem 4.3 (Convergence in $\|\cdot\|_A$-norm)**.** *With the basic assumptions stated in Section* 2.1*, assume* $\mathrm{Pe} \le 1$*, take* $\gamma = \gamma_*$*, and assume the bound* (74) *on the Courant number together with the bound* (77) *on the time step. Then,*

$$\left(\tau\sum_{n=1}^N \|u^n - u_h^n\|_A^2\right)^{1/2} \lesssim \tilde{C}_{\mathrm{tim}}t_*^{1/2}\tau + \tilde{C}_{\mathrm{spc}}\mu^{1/2}h, \tag{80}$$

eq:err.est.d

where $\tilde{C}_{\mathrm{tim}}^2 = \sum_{n=0}^{N-1}\tau(\tilde{C}_\Psi^n)^2$ *and* $\tilde{C}_{\mathrm{spc}}^2 = \sum_{n=0}^{N-1}\tau((\tilde{K}_2^n)^2 + (\tau/\mu)(K_{w-u}^n)^2)$.

Proof. Using the stability result of Lemma 4.7, we sum over n, discard the $|\cdot|_S$-terms on the left-hand side, and use a discrete Gronwall lemma to eliminate the contribution of $\|\xi_h^n\|_L^2$ in $\tilde{E}_h^n$. This yields

$$\tau\sum_{n=1}^N \|\xi_h^n\|_A^2 \lesssim \sum_{n=0}^{N-1}\tau(\|\theta_\pi^n\|_{B*}^2 + \|\zeta_\pi^n\|_{B*}^2 + \|\theta_\pi^n\|_{A*}^2 + \|\zeta_\pi^n\|_{A*}^2 + \mathrm{Pe}^{-1}(|\theta_\pi^n|_S^2 + |\zeta_\pi^n|_S^2) + t_*(\tilde{C}_\Psi^n)^2\tau^2).$$

To bound the terms with θ_π^n and ζ_π^n, we use the result of Lemma 3.7 for the $\|\cdot\|_{A*}$- and $\|\cdot\|_{B*}$-norm, while for the $|\cdot|_S$-seminorm, we use the bounds (33) and (35) on $|v^n|_{H^2}$ and $|w^n|_{H^2}$ and $|u^n|_{H^2} \le \tilde{K}_2^n$ to infer

$$\mathrm{Pe}^{-1/2}(|\theta_\pi^n|_S + |\zeta_\pi^n|_S) \lesssim \mu^{1/2}h(|v^n|_{H^2} + |w^n|_{H^2}) \lesssim \mu^{1/2}h\tilde{K}_2^n + \tau^{1/2}hK_{w-u}^n.$$

The conclusion is straightforward using $\sigma^{1/2}h^{1/2} \le \mu^{1/2}$ since $\mathrm{Pe} \le 1$. □

It is possible to derive an L-norm error estimate with higher convergence rates than (80). The proof is postponed to §7.2.

if.dom **Proposition 4.2** (Convergence in L-norm)**.** *Assume that β has bounded second-order derivatives with associated bound denoted by σ_2. Then, under the assumptions of Theorem 4.3, there holds*

$$\|u^N - u_h^N\|_L \lesssim C_{\mathrm{tim}}\tau^{3/2} + \hat{C}_{\mathrm{spc}}\sigma^{1/2}h^{3/2} + \hat{C}'_{\mathrm{spc}}\mu^{-1/2}h^2, \tag{81}$$

where C_{tim} is defined in Theorem 4.2, $(\hat{C}_{\mathrm{spc}})^2 = \sum_{n=0}^{N-1}\tau(\hat{K}_{w-u}^n)^2$, $(\hat{C}'_{\mathrm{spc}})^2 = \sum_{n=0}^{N-1}\tau C_P^2(K_2^n + \|\partial_t u\|_{C(I_n;H^2)})^2$, with $\hat{K}_{w-u}^n = C_P(|u^n|_{H^3} + (\tau/\mu)^{1/2}K_2^n + \sigma^{-1}(\sigma_1\tilde{K}_2^n + \sigma_2\tilde{K}_1^n)) + \tilde{K}_2^n + (\tau/\mu)^{1/2}K_{w-u}^n$ and C_P is the length scale associated with the Poincaré inequality stating that for all $v_h \in V_h$, $\|v_h\|_L \le \mu^{-1/2}C_P\|v_h\|_A$.

5. Numerical examples

ec:num We consider two numerical experiments using FreeFem++ [23] to illustrate the above analysis, namely convergence to a known smooth solution and control of spurious oscillations for a solution with sharp layers. For all flow regimes, we used the values $S_{\mathrm{cip}} = 0.005$ and $S_{\mathrm{bc}} = 10$ for the penalty parameters and $\gamma = 1 - \frac{1}{\sqrt{2}}$.

5.1. Convergence to smooth solutions

Let $\Omega = \{r^2 := x^2 + y^2 < 2\}$ and consider the rotating velocity field $\beta = (y, -x)^T$ so that $\sigma = 2$. Letting $\mathbf{x} = (x, y)^T$, the exact solution is chosen to be the advected heat kernel in the form

$$u(\mathbf{x}, t) = \frac{\ell_0^2}{t\mu + \ell_0^2}\exp\left(\frac{|\mathbf{r}(t) - \mathbf{x}|^2}{4(\mu t + \ell_0^2)}\right), \qquad \mathbf{r}(t) = (-0.3\sin(t), 0.3\cos(t))^T,$$

where the length scale $\ell_0 = 0.1$ determines the spread of the initial Gaussian. We consider two settings, first $\mu = 0.1$ and $t_{\mathrm{F}} = \pi/4$ and then $\mu = 10^{-4}$ and $t_{\mathrm{F}} = 2\pi$. In both cases, the decay of the exact solution away from $\mathbf{r}(t)$ is sufficiently fast to enforce homogeneous Dirichlet conditions on $\partial\Omega$. We discretize the boundary $\partial\Omega$ with M elements from which a quasi-uniform triangulation of Ω is constructed, yielding a mesh size $h = 4\pi/M$. We take $M = 2^{6+m}$ with $m \in \{0, 1, 2, 3, 4\}$. For $\mu = 0.1$, the Péclet number decays from 4 to 0.25 corresponding to a diffusion-dominated regime, while for $\mu = 10^{-4}$, the Péclet number is 10^3 times larger, corresponding to an advection-dominated regime. In both regimes, the time step is selected by setting the Courant number to $\mathrm{Co} = \frac{1}{2}$. Results are reported in Table 5.1. For $\mu = 0.1$, the result on the finest mesh is omitted since the mesh is sufficiently fine, and the diffusion coefficient sufficiently large, to detect the influence of using homogeneous Dirichlet boundary conditions; for $\mu = 10^{-4}$, the result on the coarsest mesh is omitted since the mesh is too coarse to resolve the initial datum. In all cases, the convergence rates match, or are slightly better than, those predicted by the theory.

5.2. Solutions with sharp layers

The purpose of this test case is to illustrate numerically that in the advection-dominated regime, spurious oscillations resulting from insufficient mesh resolution of sharp layers do not spread over the whole domain, but

smooth

m	$\mu = 0.1$		$\mu = 10^{-4}$
	$L_t^\infty(L^2)$	$L_t^2(H^1)$	$L_t^\infty(L^2)$
0	2.8e-3	5.7e-2	—
1	9.5e-4	2.7e-2	4.6e-2
2	2.2e-4	1.0e-2	1.0e-2
3	5.3e-5	5.5e-3	1.4e-3
4	—	—	2.1e-4

TABLE 1. Convergence for smooth solution

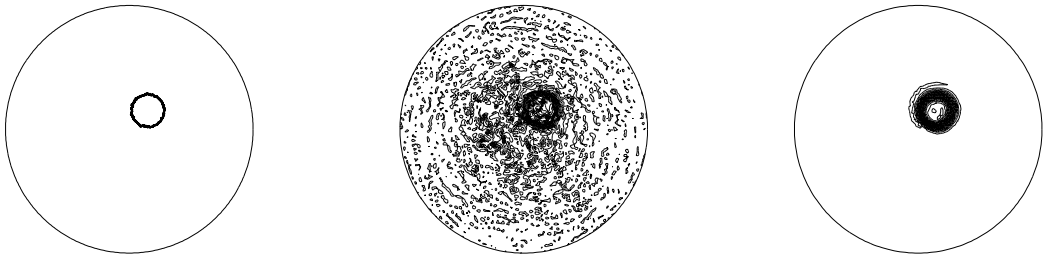

FIGURE 1. Initial data (left) and solution at final time without (middle) and with (right) stabilization

rough_adv

remain contained at all times close to the layer. Let $\mu = 10^{-6}$ and consider the initial datum

$$u_0(\mathbf{x}) = 0.5(\tanh((\exp(-20|\mathbf{r}(\tfrac{\pi}{4}) - \mathbf{x}|^2) - 0.5)/0.0001) + 1).$$

The graph of u_0 corresponds to a cylinder centered at $\mathbf{x} = (0.3, 0.3)^T$. The width of the inner layer is 10^{-4}. The mesh is built using $M = 512$ so that it does not resolve this inner layer. The final time is $t_{\mathrm{F}} = 2\pi$ corresponding to one full rotation of the initial datum. Figure 1 displays the initial datum, the approximate solution without stabilization ($S_{\mathrm{cip}} = 0$), and the solution with stabilization ($S_{\mathrm{cip}} = 0.005$). The unstabilized solution exhibits global spurious oscillations, while the improved quality of the stabilized solution is clearly visible.

6. EXTENSIONS

ec:ext

For simplicity, the above analysis was presented in the case where space discretization was performed using continuous, piecewise affine finite elements with CIP. Other finite element methods with symmetric stabilization can be used. This requires establishing discrete stability and boundedness for the discrete operators B_h and A_h. For consistent methods, the stability and convergence analysis of §4 can then be readily applied, while minor adaptations are needed in the case of nonconsistent methods to formulate the truncation errors.

To illustrate, we briefly consider a DG method for space discretization using upwinding for the advective part and symmetric interior penalty for the diffusive part. Let V_h^{d} denote the space spanned by (discontinuous) piecewise affine functions on the mesh $\mathcal{T}_h$. For a smooth enough function v that is possibly double-valued at $F \in \mathcal{F}_h^{\mathrm{int}}$ with $F = \partial T^- \cap \partial T^+$, we define, in addition to its jump, its mean value as $\{\!\{v\}\!\} := \frac{1}{2}(v|_{T^-} + v|_{T^+})$. On boundary faces, the jump and mean value refer to the actual value of v on F. The discrete operators B_h

and A_h are now such that

$$(B_h z, w_h)_L := (\beta\cdot\nabla_h z, w_h)_L - \sum_{F\in\mathcal{F}_h^{\rm int}} ((\nu_F\cdot\beta)[\![z]\!], \{\!\{w_h\}\!\})_{L,F} + \sum_{F\in\mathcal{F}_h^{\rm int}} S_{\rm upw}(|\nu_F\cdot\beta|[\![z]\!], [\![w_h]\!])_{L,F},$$

$$(A_h z, w_h)_L = \mu(\nabla_h z, \nabla_h w_h)_{L^d} - \sum_{F\in\mathcal{F}_h} \mu(\nu_F\cdot\{\!\{\nabla_h z\}\!\}, [\![w_h]\!])_{L,F} - \sum_{F\in\mathcal{F}_h} \mu([\![z]\!], \nu_F\cdot\{\!\{\nabla_h w_h\}\!\})_{L,F}$$
$$+ \sum_{F\in\mathcal{F}_h} S_{\rm ip}\mu h_F^{-1}([\![z]\!], [\![w_h]\!])_{L,F},$$

where ∇_h denotes the broken (elementwise) gradient operator, while $S_{\rm upw} = \frac{1}{2}$ for classical upwinding, and $S_{\rm ip}$ is taken large enough. Then, letting

$$|z|_S^2 := \sum_{F\in\mathcal{F}_h^{\rm int}} \tfrac{1}{2}\||\nu_F\cdot\beta|^{1/2}[\![z]\!]\|_{L,F}^2, \qquad \|z\|_A^2 := \mu\|\nabla z\|_{L^d}^2 + \sum_{F\in\mathcal{F}_h} \mu h_F^{-1}\|[\![z]\!]\|_{L,F}^2,$$

it is readily verified that the discrete stability properties stated in §2.6 hold true. Moreover, letting

$$\|z\|_{B*} := |z|_S + \sigma^{1/2}h^{-1/2}\|z\|_L + \left(\sum_{T\in\mathcal{T}_h}\sum_{F\subset\partial T} \sigma\|z\|_{L,F}^2\right)^{1/2},$$

$$\|z\|_{A*} := \|z\|_A + \left(\sum_{T\in\mathcal{T}_h}\sum_{F\subset\partial T} \mu h_F\|\nu_F\cdot\nabla z\|_{L,F}^2\right)^{1/2},$$

it is readily verified that the boundedness properties stated in §2.6 hold true.

Finally, it is also possible to consider higher-order continuous or discontinuous finite elements with symmetric stabilization. To achieve stability, the sole modification in the above analysis concerns the advection-dominated regime, since the discrete inverse inequality (61) can no longer be used. It is then necessary to modify the proof of Lemma 4.3 when bounding $\frac{1}{2}\|\xi_h^{n+1} - \zeta_h^n\|_L^2$. In particular, following [8] (details are skipped for brevity), the term $\frac{1}{2}\tau B_h\eta_h^n$ on the right-hand side of (64) is controlled by the so-called 4/3-CFL condition $\tau \lesssim t_*^{-1/3}(h/\sigma)^{4/3}$. Deriving convergence rates is a more delicate question not covered herein; it entails, in particular, obtaining bounds for higher-order Sobolev norms of the auxiliary functions v^n and w^n.

7. Proofs of Propositions 4.1 and 4.2

proofs

This sections collects the proofs of Propositions 4.1 and 4.2.

7.1. **Proof of Proposition 4.1**

fs.adv

The proof, which proceeds along that of Lemma 4.7, is only sketched. There are essentially two differences. Firstly, the term T_2 in this proof needs to be bounded since we do not assume here that $\gamma = \gamma_*$. To this purpose, we use (16a) and the definition of α_h^n to obtain

$$\tau(\xi_h^n, A_h\xi_h^{n+1})_L = \tau(\theta_h^n, A_h\xi_h^{n+1})_L + \gamma\tau^2(A_h\theta_h^n, A_h\xi_h^{n+1})_L - \gamma\tau^2(A_h\theta_\pi^n, A_h\xi_h^{n+1})_L.$$

The first term on the right-hand side is treated as the term T_1 in the proof of Lemma 4.7. For the second term, the Cauchy–Schwarz inequality and (60) yield

$$\tau^2(A_h\theta_h^n, A_h\xi_h^{n+1})_L \le \tau^2\|A_h\theta_h^n\|_L\|A_h\xi_h^{n+1}\|_L \lesssim ({\rm Co/Pe})\tau\|\theta_h^n\|_A\|\xi_h^{n+1}\|_A \le \tau\|\theta_h^n\|_A\|\xi_h^{n+1}\|_A,$$

since $\mathrm{Co} \le 1$ and $\mathrm{Pe} \ge 1$. Finally, for the third term, the Cauchy–Schwarz inequality, (60), and (26) lead to

$$\tau^2(A_h\theta_\pi^n, A_h\xi_h^{n+1})_L \le \tau^2\|A_h\theta_\pi^n\|_L\|A_h\xi_h^{n+1}\|_L \lesssim \tau\|\theta_\pi^n\|_{A*}\|\xi_h^{n+1}\|_A,$$

since $\tau^{1/2}\mu^{1/2}h^{-1} = (\mathrm{Co}/\mathrm{Pe})^{1/2} \le 1$. Collecting these estimates, we infer

$$\tau(T_2, A_h\xi_h^{n+1})_L \lesssim \tau(\|\theta_h^n\|_A + \|\theta_\pi^n\|_{A*})\|\xi_h^{n+1}\|_A.$$

Secondly, when dealing with the truncation error in time, we exploit the fact that $\mathrm{Pe} \ge 1$ to derive a sharper bound than in the proof of Lemma 4.7, namely

$$\tau^2(\Psi_h^n, A_h\xi_h^{n+1})_L \le \tau^2\|\Psi_h^n\|_L\mu^{1/2}h^{-1}\|\xi_h^{n+1}\|_A \le (\mathrm{Co}/\mathrm{Pe})^{1/2}\tau^{3/2}\|\Psi^n\|_L\|\xi_h^{n+1}\|_A \lesssim \tau E_h^n\|\xi_h^{n+1}\|_A,$$

where we have used (66). As a result, an estimate similar to (78) is inferred, but with a quantity $\hat{E}_h^n$ on the right-hand side which is defined as (76) except that $t_*^{1/2}\tilde{C}_\Psi^n\tau$ is replaced by the sharper estimate $C_\Psi^n\tau^{3/2}$. The conclusion is straightforward using, in particular, that

$$\mathrm{Pe}^{-1/2}(|\theta_\pi^n|_S + |\zeta_\pi^n|_S) \lesssim \mu^{1/2}h(|v^n|_{H^2} + |w^n|_{H^2}) \lesssim \mu^{1/2}h\tilde{K}_2^n + \tau^{1/2}hK_{w-u}^n \le \sigma^{1/2}h^{3/2}(\tilde{K}_2^n + \sigma^{-1}K_{w-u}^n),$$

since $\mathrm{Pe} \ge 1$ and $\mathrm{Co} \le 1$.

fs.dif

7.2. **Proof of Proposition 4.2**

For brevity, we only sketch the proof. We introduce the discrete Riesz projection of u^n and of the auxiliary functions v^n and w^n. Specifically, $r_hu^n \in V_h$ is defined such that $A_hr_hu^n := A_hu^n$ and similarly for r_hv^n and r_hw^n. Then, redefining the quantities $\xi_h^n := u_h^n - r_hu^n$, $\xi_\pi^n := u^n - r_hu^n$ and similarly for θ_h^n, θ_π^n, ζ_h^n, and ζ_π^n, the error equation takes again the form (16) with the new source terms

$$\alpha_h^n = \tau^{-1}\pi_h(I - r_h)(v^n - u^n), \qquad \beta_h^n = \tau^{-1}\pi_h(I - r_h)(w^n - v^n) - B_h(I - r_h)v^n,$$
$$\delta_h^n = \tau^{-1}\pi_h(I - r_h)(u^{n+1} - \tfrac{1}{2}(v^n + w^n)) - \tfrac{1}{2}B_h(I - r_h)w^n.$$

Then, the basic energy identity of Lemma 4.1 is not modified. Instead, the basic energy estimate of Theorem 4.1 requires bounding the new source terms. Using the Cauchy–Schwarz inequality, the Poincaré inequality, the approximation properties of the Riesz projector r_h, the bound (32) on $\|\Delta(v^n - u^n)\|_L$, and elliptic regularity, we obtain

$$\tau(\alpha_h^n, \theta_h^n)_L \le \mu^{-1/2}C_P\|(I - r_h)(v^n - u^n)\|_L\|\theta_h^n\|_A \lesssim \mu^{-1/2}C_Ph^2|v^n - u^n|_{H^2}\|\theta_h^n\|_A \lesssim \mu^{-1/2}C_Ph^2\tau K_2^n\|\theta_h^n\|_A.$$

Hence, by Young's inequality,

$$\tau(\alpha_h^n, \theta_h^n)_L \le C\tau(\mu^{-1/2}C_Ph^2K_2^n)^2 + \lambda\tau\|\theta_h^n\|_A^2,$$

where λ can be chosen as small as needed. To bound $\tau(\beta_h^n, \theta_h^n)_L$, we write $w^n - v^n = (w^n - u^n) - (v^n - u^n)$, and estimate the contribution of $(v^n - u^n)$ as for α_h^n. To bound the contribution of $(w^n - u^n)$, we observe that

$$\|(I - r_h)(w^n - u^n)\|_L \lesssim h^2|w^n - u^n|_{H^2} \lesssim h^2\|\Delta(w^n - u^n)\|_L.$$

We use a different bound on $\|\Delta(w^n - u^n)\|_L$ than (34), whereby we exploit that the advection field β has bounded second-order derivatives. Letting $\mathfrak{v}$ denote the right-hand side of (36) and observing that $\mathfrak{v} \in V$, (31) yields $\|\Delta(w^n - u^n)\|_L \lesssim \|\Delta\mathfrak{v}\|_L$. Using the bounds (33) on v^n and the bound (32) on $\|\nabla\Delta(v^n - u^n)\|_{L^d}$, we

infer $\|\Delta(Bv^n)\|_L \lesssim \sigma\tilde{K}^n_{w-u}$ with $\tilde{K}^n_{w-u} = |u^n|_{H^3} + (\tau/\mu)^{1/2}K^n_2 + \sigma^{-1}(\sigma_1\tilde{K}^n_2 + \sigma_2\tilde{K}^n_1)$. Hence, $\|\Delta(w^n - u^n)\|_L \lesssim \tau(K^n_2 + \sigma\tilde{K}^n_{w-u})$, so that

$$\|(I - r_h)(w^n - u^n)\|_L \lesssim \tau h^2(K^n_2 + \sigma\tilde{K}^n_{w-u}).$$

Finally, for the last term in β^n_h, we obtain by integrating by parts the advective derivative that

$$(B_h(I - r_h)v^n, \theta^n_h)_L \lesssim h^2|v^n|_{H^2}\sigma\mu^{-1/2}\|\theta^n_h\|_A \lesssim \sigma\mu^{-1/2}h^2\tilde{K}^n_2\|\theta^n_h\|_A,$$

since $|v^n|_{H^2} \lesssim \tilde{K}^n_2$. Collecting these bounds and introducing the Péclet number yields

$$\tau(\beta^n_h, \theta^n_h)_L \le C\tau(\mu^{-1/2}h^2C_PK^n_2 + \mathrm{Pe}^{1/2}\sigma^{1/2}h^{3/2}(C_P\tilde{K}^n_{w-u} + \tilde{K}^n_2))^2 + \lambda\tau\|\theta^n_h\|^2_A,$$

where λ can be chosen as small as needed. The bound on $\tau(\delta^n_h, \zeta^n_h)_L$ is obtained using similar arguments, in particular that $u^{n+1} - \frac{1}{2}(v^n + w^n) = (u^{n+1} - u^n) - (\frac{1}{2}(v^n + w^n) - u^n)$, $\|(I - r_h)(u^{n+1} - u^n)\|_L \lesssim \tau h^2\|\partial_t u\|_{C(I_n;H^2)}$, and that $|w^n|_{H^2} \lesssim \tilde{K}^n_2 + (\tau/\mu)^{1/2}K^n_{w-u}$ owing to (35). Therefore, we recover the stability estimates (59) and (70) with

$$E^n_h = \mu^{-1/2}h^2C_P(K^n_2 + \|\partial_t u\|_{C(I_n;H^2)}) + \mathrm{Pe}^{1/2}\sigma^{1/2}h^{3/2}\hat{K}^n_{w-u}, \tag{82}$$

eq:E.L.Pe

with $\hat{K}^n_{w-u} = C_P\tilde{K}^n_{w-u} + \tilde{K}^n_2 + (\tau/\mu)^{1/2}K^n_{w-u}$. The next step is the result of Lemma 4.5 where the identity (72) holds true with

$$\Xi^n_h = \omega_2\alpha^n_h + (\omega_1 + \tfrac{1}{2})\beta^n_h - \delta^n_h.$$

Then, proceeding as in Lemma 4.6, we need to control $\tau\|\Xi^n_h\|_L$. We observe that

$$\tau\|B_h(I - r_h)v^n\|_L + \tau\|B_h(I - r_h)w^n\|_L \lesssim \tau\sigma h(|v^n|_{H^2} + |w^n|_{H^2}) \lesssim \tau^{1/2}\sigma^{1/2}h^{3/2}(\tilde{K}^n_2 + (\tau/\mu)^{1/2}K^n_{w-u}).$$

Defining $\hat{E}^n_h$ as E^n_h in (82) by dropping the $\mathrm{Pe}^{1/2}$ factor in the last term, that is,

$$\hat{E}^n_h = \mu^{-1/2}h^2C_P(K^n_2 + \|\partial_t u\|_{C(I_n;H^2)}) + \sigma^{1/2}h^{3/2}\hat{K}^n_{w-u},$$

we eventually infer $\tau\|\Xi^n_h\|_L \lesssim \tau^{1/2}\hat{E}^n_h$. Finally, accounting for the truncation error in time, we recover the stability estimate (75) with the right-hand side

$$\bar{E}^n_h := \mu^{-1/2}h^2C_P(K^n_2 + \|\partial_t u\|_{C(I_n;H^2)}) + \sigma^{1/2}h^{3/2}\hat{K}^n_{w-u} + C^n_\Psi\tau^{3/2},$$

whence the conclusion is straightforward.

Remark 7.1 (Optimality in h). We observe that the error term defined by (82) exhibits second-order convergence as $h \to 0$ owing to the presence of the $\mathrm{Pe}^{1/2}$ factor in the last term. This is no longer the case for the error term $\bar{E}^n_h$, where the loss of the $\mathrm{Pe}^{1/2}$ factor is caused by the contribution of B_h when bounding the anti-dissipative term. Optimality is recovered for vanishing advection.

References

ARS:97 [1] U. M. Ascher, S. J. Ruuth, and R. J. Spiteri. Implicit-explicit Runge-Kutta methods for time-dependent partial differential equations. *Appl. Numer. Math.*, 25(2-3):151–167, 1997. Special issue on time integration (Amsterdam, 1996).

ARW:95 [2] U. M. Ascher, S. J. Ruuth, and B. T. R. Wetton. Implicit-explicit methods for time-dependent partial differential equations. *SIAM J. Numer. Anal.*, 32(3):797–823, 1995.

BJL.07 [3] M. Braack, E. Burman, V. John, and G. Lube. Stabilized finite element methods for the generalized Oseen problem. *Comput. Methods Appl. Mech. Engrg.*, 196(4-6):853–866, 2007.

BH:82 [4] A. N. Brooks and T. J. R. Hughes. Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations. *Comput. Methods Appl. Mech. Engrg.*, 32(1-3):199–259, 1982. FENOMECH '81, Part I (Stuttgart, 1981).

Bu05 [5] E. Burman. A unified analysis for conforming and nonconforming stabilized finite element methods using interior penalty. *SIAM J. Numer. Anal.*, 43(5):2012–2033 (electronic), 2005.

Bu10 [6] E. Burman. Consistent SUPG-method for transient transport problems: Stability and convergence. *Comput. Methods Appl. Mech. Engrg.*, 199(17-20):1114–1123, 2010.

7_m2an [7] E. Burman and A. Ern. A continuous finite element method with face penalty to approximate Friedrichs' systems. *M2AN Math. Model. Numer. Anal.*, 41(1):55–76, 2007.

BEF:10 [8] E. Burman, A. Ern, and M. A. Fernández. Explicit Runge–Kutta schemes and finite elements with symmetric stabilization for first-order linear PDE systems. *SIAM J. Numer. Anal.*, 2010. To appear, `http://hal.archives-ouvertes.fr/hal-00380659`.

BF09 [9] E. Burman and M. A. Fernández. Finite element methods with symmetric stabilization for the transient convection-diffusion-reaction equation. *Comput. Methods Appl. Mech. Engrg.*, 198(33-36):2508–2519, 2009.

BGL:09 [10] E. Burman, J. Guzmán, and D. Leykekhman. Weighted error estimates of the continuous interior penalty method for singularly perturbed problems. *IMA J. Numer. Anal.*, 29(2):284–314, 2009.

BH04 [11] E. Burman and P. Hansbo. Edge stabilization for Galerkin approximations of convection–diffusion–reaction problems. *Comput. Methods Appl. Mech. Engrg.*, 193:1437–1453, 2004.

BS10 [12] E. Burman and G. Smith. Analysis of the space semi-discretized SUPG method for transient convection–diffusion equations. Technical report, University of Sussex, 2010.

cSh:89 [13] B. Cockburn and C.-W. Shu. TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws. II. General framework. *Math. Comp.*, 52(186):411–435, 1989.

Cod00 [14] R. Codina. Stabilization of incompressibility and convection through orthogonal sub-scales in finite element methods. *Comput. Methods Appl. Mech. Engrg.*, 190(13-14):1579–1599, 2000.

Cod02 [15] R. Codina. Stabilized finite element approximation of transient incompressible flows using orthogonal subscales. *Comput. Methods Appl. Mech. Engrg.*, 191(39-40):4295–4321, 2002.

Cr80 [16] M. Crouzeix. Une méthode multipas implicite-explicite pour l'approximation des équations d'évolution paraboliques. *Numer. Math.*, 35(3):257–276, 1980.

DEG08 [17] D. A. Di Pietro, A. Ern, and J.-L. Guermond. Discontinuous Galerkin methods for anisotropic semidefinite diffusion with advection. *SIAM J. Numer. Anal.*, 46(2):805–831, 2008.

EG04 [18] A. Ern and J.-L. Guermond. *Theory and Practice of Finite Elements*, volume 159 of *Applied Mathematical Sciences*. Springer-Verlag, New York, NY, 2004.

richs1 [19] A. Ern and J.-L. Guermond. Discontinuous Galerkin methods for Friedrichs' systems. I. General theory. *SIAM J. Numer. Anal.*, 44(2):753–778, 2006.

Gu99 [20] J.-L. Guermond. Stabilization of Galerkin approximations of transport equations by subgrid modeling. *Math. Model. Numer. Anal. (M2AN)*, 33(6):1293–1316, 1999.

Gu01 [21] J.-L. Guermond. Subgrid stabilization of Galerkin approximations of linear monotone operators. *IMA J. Numer. Anal.*, 21:165–197, 2001.

Gu:06 [22] J. Guzmán. Local analysis of discontinuous Galerkin methods applied to singularly perturbed problems. *J. Numer. Math.*, 14(1):41–56, 2006.

reefem [23] F. Hecht. *FreeFem++, Third Edition, Version 3.0-1. User's Manual*. LJLL, University Paris VI.

JNP:84 [24] C. Johnson, U. Nävert, and J. Pitkäranta. Finite element methods for linear hyperbolic problems. *Comput. Methods Appl. Mech. Engrg.*, 45(1-3):285–312, 1984.

JP86 [25] C. Johnson and J. Pitkäranta. An analysis of the discontinuous Galerkin method for a scalar hyperbolic equation. *Math. Comp.*, 46(173):1–26, 1986.

LR74 [26] P. Lesaint and P.-A. Raviart. On a finite element method for solving the neutron transport equation. In C. de Boors, editor, *Mathematical aspects of Finite Elements in Partial Differential Equations*, pages 89–123. Academic Press, 1974.

Tadmor [27] D. Levy and E. Tadmor. From semidiscrete to fully discrete: stability of Runge-Kutta schemes by the energy method. *SIAM Rev.*, 40(1):40–73 (electronic), 1998.

PR:05 [28] L. Pareschi and G. Russo. Implicit-Explicit Runge-Kutta schemes and applications to hyperbolic systems with relaxation. *J. Sci. Comput.*, 25(1-2):129–155, 2005.

ook.08 [29] H.-G. Roos, M. Stynes, and L. Tobiska. *Robust numerical methods for singularly perturbed differential equations*, volume 24 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2008. Convection-diffusion-reaction and flow problems.